NELSON'S AEROSCIENCE MANUALS

GENERAL EDITOR:
H. LEVY, M.A., D.Sc., F.R.S.E.

SUB-ATOMIC PHYSICS

This book forms Volume II of a course of PHYSICS FOR AERONAUTICAL STUDENTS, complete in two volumes, by Professor Dingle.

Volume I is entitled *Mechanical Physics*.

SUB-ATOMIC PHYSICS

by

HERBERT DINGLE

Professor of Natural Philosophy, Imperial College of Science and Technology (University of London)

Revised Edition

THOMAS NELSON AND SONS LTD
LONDON EDINBURGH PARIS MELBOURNE
TORONTO AND NEW YORK

First published September 1942
Reprinted 1942, 1944
Enlarged and revised edition 1946

PREFACE

THIS text-book is designed to present a general view of modern physics in a form suited to the needs of students of aeronautics, though it is hoped the book will also prove helpful to readers who have any of the ordinary examinations in view. At the same time, it is essentially a text-book of physics and not of aeronautics. What has been aimed at is the exposition of physical principles in a manner that will enable their application to aeronautical and related studies to be readily understood. Hence attention has been concentrated on giving a clear account of the ideas of the subject and of the results of physical research. In short, the book aims at giving a vivid general picture of the present state of physical knowledge and a practical understanding of what it means.

Physics is now such an enormous subject that comprehensiveness can be attained only at considerable cost of much that it would be desirable to achieve. It is impossible, for instance, to make the book adequate as a guide to laboratory work, and it has been thought undesirable to try to do at all what cannot be done sufficiently well.

No previous knowledge of physics is necessary to the understanding of the book, but the general attainments of the reader are assumed to be those of the post-matriculation stage. Some knowledge of elementary mechanics is essential. Mathematics has been reduced to a minimum, but the Calculus has not been avoided when much insight into physical principles can be obtained by its use. Nevertheless, it makes only rare appearances, and the student unacquainted with this branch of mathematics may omit without serious loss those sections in which it appears.

It has been my intention to make the book suitable for any of the training schemes for supplying the technical needs of the Services. The more elementary students will therefore find some portions beyond them, and the guidance of the teacher is very desirable in enabling them to make the best use of the book. It will be found, however, that knowledge of the more difficult portions has not been

presumed in what follows, and the student's own attainments may be a sufficient guide to what he may pass over with safety.

Finally, a word should be said on the order of treatment adopted. The main division is between what I have called "mechanical physics" and "sub-atomic physics." At the present time it is very convenient to distinguish those departments in which the atom or molecule can still be treated as an elementary particle from those in which the structure of the atom is fundamental. Mechanical physics and sub-atomic physics form two separate volumes in this series.

The effect of this classification on the traditional division of the subject into Heat, Light, Sound, Properties of Matter, Magnetism, and Electricity, is broadly to place Heat, Sound, and Properties of Matter under Mechanical Physics, and Light, Magnetism, and Electricity under Sub-atomic Physics; but the older division is, in any case, obsolete, since many of the most vital of modern physical studies—such as radioactivity, X-rays, photo-electricity, etc.—had no place in it. Little trace of it remains here, for in each volume the aim has been to show the connection rather than the boundaries between the various physical phenomena.

It only remains to add that Figs. 146, 147, 148, and 149 are reproduced from *The Earth's Magnetism* by Professor S. Chapman (Methuen's "Monographs on Physical Subjects"), to whom and to the publishers the author's thanks are due.

HERBERT DINGLE

IMPERIAL COLLEGE
January 1942

PREFACE TO THE REVISED EDITION

In this edition a few errors have been corrected: I wish to take this opportunity of expressing my thanks to correspondents who have called my attention to them. Since the book was first written, information concerning the atomic bomb has been released, and an account of the general question of the utilization of nuclear energy, with particular reference to this weapon, has been added to Chapter XIV.

H. D.

***September* 1945**

CONTENTS

PART I.—INTRODUCTORY

PART II.—OPTICS

PART III.—ELECTRICITY AND MAGNETISM

References to *Mechanical Physics*, the first volume of this course, are indicated by the symbol I, followed by the page number.

SUB-ATOMIC PHYSICS

PART I

INTRODUCTORY

CHAPTER I

THE STRUCTURE OF THE ATOM

In the first volume of this book we have, for the most part, treated matter as though it were entirely inert, and studied its behaviour under the influence of external forces. Only occasionally have we met with indications that it is itself sometimes the origin of forces, and an active instead of a passive constituent of the world. It is true that the universal force of gravitation suggests a certain activity in matter, but, except when bodies of astronomical size are concerned, this is negligible and has been omitted from our considerations. Only in such phenomena as those of surface tension, latent heat, and internal work have we been led to consider the atoms of matter as exerting an influence on other atoms, and it has not been necessary even here, in order to explain the facts observed, to go beyond the simple assumption that atoms attract one another with a force which decreases as they get farther apart.

In the present volume we take up the study of facts which force us to attribute activity to matter, and in order to explain this we find ourselves compelled to take up the question of the structure of the atom. The activity of matter must be attributed ultimately to the activity of atoms, and its particular character indicates that the atom must be

constructed in a particular way. Let us look at some of the phenomena which lead us to this conclusion.

Electrification

Most people are aware that if a piece of ebonite (*e.g.* the case of a fountain pen) is rubbed against the coat sleeve it will attract light bodies, such as small pieces of paper. This behaviour is not confined to ebonite. It is shown by a large number of substances, and is, in fact, a fairly general property of matter, provided that the two bodies rubbed together are of different materials. Bodies which show it are said to be *electrified*, or *charged with electricity*.

It is easily seen that we are dealing here with a force enormously greater than that of gravitation between the bodies concerned. For example, in a particular experiment a piece of ebonite weighing $2\frac{1}{2}$ gm. easily lifted a piece of paper weighing 0·002 gm. from a distance of 1 mm. The attractive force therefore overcame a weight of $0{\cdot}002 \times 981 = 1{\cdot}96$ dynes. The gravitational attraction between the ebonite and paper, however, even supposing the whole weight of each body was concentrated at the point of nearest approach, would be $G \times \dfrac{2{\cdot}5 \times 0{\cdot}002}{0{\cdot}1^2}$ dynes, where $G = 6{\cdot}6 \times 10^{-8}$, and this is equal to $3{\cdot}3 \times 10^{-8}$ dynes. The electrical attraction was therefore some 60 million times a greatly over-estimated gravitational attraction in this case.

On examination, electrification turns out to be a less simple phenomenon than gravitation. If a light piece of rubbed ebonite is suspended by a dry silk thread, and a second piece of rubbed ebonite is held near it, the first is seen to be repelled from the second. In fact, we find generally that *any* electrified body tends to repel a similarly electrified body. On the other hand, if the flannel with which the ebonite was rubbed is held near it there is attraction between them. Hence the same piece of matter can attract some bodies and repel others. Gravitation, however (phenomena occurring in

distant parts of the universe being ignored), is always an attractive force.

We shall consider the properties of electrified bodies in more detail later. At the moment we are simply concerned with the fact that a strong force of attraction or repulsion can exist between pieces of matter, and this we have to explain if we can.

It may be said at once that we cannot "explain" it in the ordinary sense of the word; at present, at least, we have to take it as an ultimate fact. But what we can do is to form an idea of the structure of the atom which will enable us to describe the electrical properties of matter in terms of atomic processes, and so discover connections between facts which otherwise would appear independent. We do this in the following way.

Positive and Negative Electrification: We first of all distinguish the attracting and repelling properties of electrified bodies by assuming that there are two kinds of electrification, which we call *positive* and *negative* respectively. A positively electrified body repels another positively electrified body, and a negatively electrified body repels another negatively electrified body, but a positively and a negatively electrified body attract each other. When two bodies are rubbed together, and so electrified, they always acquire opposite kinds of electrification, but which is of the kind arbitrarily called "positive," and which "negative," depends on the chemical nature of the bodies. Any electrified body, however, when held near a neutral (*i.e.* unelectrified) body, "induces" the opposite kind of electrification in it; this accounts, for example, for the attraction between rubbed ebonite and paper—an attraction which is exerted also by the rubbed flannel on a similar piece of paper.

The terms "positive" and "negative" are justified by the fact that when oppositely charged bodies are brought into contact the resultant charge, or amount of electrification

(the method of measuring electrification will be described later), is equal to the algebraical sum of the original charges. For example, the opposite charges produced on the flannel and ebonite are numerically equal to one another, and when the two bodies are brought into contact again they become unelectrified—the two charges neutralizing one another.

Sub-atomic Particles

We next assume that the atoms of all bodies are constructed from three kinds of particles—one positively electrified (the *proton*), one negatively electrified (the *electron*), and one unelectrified (the *neutron*). There are the same number of protons as there are electrons in every normal atom, and the amount of electrification is the same for all particles of both kinds. The result is that the total amount of electrification in the atom is zero, for the neutrons, of course, contribute no electrification. This is true of the atoms of every element, and the various chemical elements are distinguished from one another simply by the number of particles and the way in which they are arranged.

Before describing the various arrangements, we may remark that this idea of the constituents of atoms is not sheer guesswork. Particles of all three kinds have actually been separated from atoms and detected in the free state. Their charges and also their masses have been measured. The mass of the proton, as well as that of the neutron, is about the same as the mass of a single hydrogen atom, but the mass of the electron is only about $\frac{1}{1840}$ of this. The ordinary hydrogen atom (the simplest of all atoms) appears, in fact, to consist merely of one proton and one electron.

It is possible to regard the neutron as a proton and an electron united, their charges neutralizing one another, and the combined mass being so close to that of the proton alone as to be indistinguishable by present means of measurement. Alternatively, we may regard the proton as a neutron positively electrified by the loss of an electron. For certain

reasons, however, neither of these views has found favour, and the present belief is that the proton, neutron, and electron are independent entities. We regard the charge as an essential characteristic of both the proton and the electron, so that it would be impossible to reduce either of these particles to an uncharged condition. We attempt, in fact, to explain all the phenomena shown by ordinary electrified bodies as the result of an excess or defect of protons or electrons, and this explanation clearly cannot be applied to the individual particles themselves. For this reason protons and electrons are sometimes called particles of electricity, rather than electrified particles, and the term is a good one in so far as it serves to distinguish the casual and destructible electrification of ordinary bodies from the intrinsic, essential electrical character of the particles.

Atomic Structure

Let us now see how atoms are built up of these particles. In the first place, we have the protons and neutrons associated together in a close compact aggregation which we call the *nucleus* of the atom, and round this nucleus the electrons travel in orbits, something like the orbits of the planets round the Sun. This is true of all atoms, of whatever element. The differences between the various elements consist simply in the number of protons in the nucleus—and therefore also in the number of satellite electrons, since, as we have seen, this is the same as the number of protons.

Let us begin with the lightest element, hydrogen. The atom of hydrogen, as we have said, contains simply one proton and one electron. The proton forms the nucleus, and the electron travels round it. The next element, helium, has a nucleus consisting of two protons and two neutrons, and round this, of course, two electrons must circulate. Lithium, the third element, has three protons and four neutrons in its nucleus, and three revolving electrons. As we go through the periodic table of elements we find that

at each step we add one nuclear proton and one revolving electron, and usually a neutron or two, in order to obtain the next element. The number of protons (or electrons) in the atom is called the *atomic number* of the element. It ranges from 1 for hydrogen to 92 for uranium, with only a few gaps now remaining to be filled by undiscovered elements. Fig. 1 shows, in purely diagrammatic form, the atom of helium.

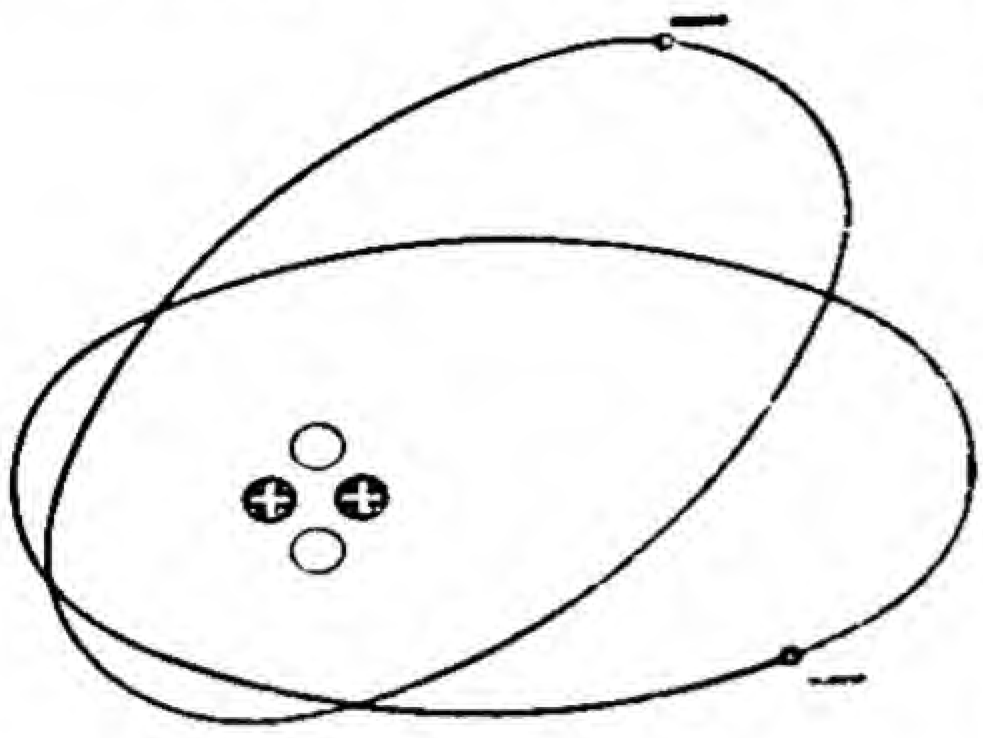


FIG. 1
Model of the Helium Atom
The nucleus contains two (positive) protons and two neutrons, round which circulate two (negative) electrons in separate orbits

Isotopes

In this scheme it will be noticed that the number of protons or electrons makes all the difference between one element and another, but the number of neutrons seems to have no special significance. That appears to be the fact. An additional neutron or two in the nucleus makes no difference to the chemical properties of the atom ; it simply adds to the atomic weight. We know, in fact, that the observable samples of most elements consist of a mixture of atoms of differing weight, and therefore differing neutron content. Such different forms of the same element are called *isotopes*. Hydrogen, for example, has two isotopes, the one we have described being by far the commoner. The other has a neutron as well as a proton in its nucleus, and its weight is therefore about twice that of the ordinary hydrogen atom. It is very remarkable that, with very few exceptions, wherever we find hydrogen (or indeed, any other element which has isotopes) on the Earth the isotopes are found in it in the same proportion. With hydrogen, the lighter isotope is several thousand times more abundant than the heavier. With chlorine (atomic number 17), the two isotopes, of atomic

weights 35 and 37, are present in the ratio 3 : 1, so that the experimentally determined atomic weight of chlorine is 35·5. It is this constancy in the proportions with which the isotopes are mixed that accounts for the agreement between the measurements of atomic weight (which are actually measurements of the mean atomic weight of the several isotopes) of different samples of an element.

We shall not need to consider isotopes further, nor in most of our work shall we need to concern ourselves with the structure of the nucleus of an atom. We may regard the aggregation of protons and neutrons as a single positive charge of value $+Z$, where Z is the number of protons in it, and the unit of charge is the charge of a single proton. The number of revolving electrons is then also Z, and this number is the atomic number of the element in question. This is the picture of the atom which we must see in our mind's eye in considering the phenomena of electricity.

Explanation of Electrification

We can now give a rough explanation of the production of electrification by rubbing. Friction between the atoms of the two substances concerned results in the detachment of electrons—which, being the outermost parts of the atoms and the least massive, are the most readily detached. Now the ease with which electrons are broken away is different for different elements, so that, in general, when two different substances are rubbed together more electrons are set free in one than in the other. Some of them wander over into the other substance, and the result is that that substance acquires an excess of negative charge, while the first is left with an equal excess of positive charge.

The energy needed to remove an electron from an atom is clearly an important quantity, for it determines which of the two substances will be positively and which negatively charged. We may give a partial account of the matter by

describing in a little more detail how the electrons are arranged round the nucleus.

Arrangement of Atomic Electrons

The arrangement is in groups, or *shells*—to use the customary name. Consider, for example, an element of fairly high atomic number which has many electrons. The two nearest to the nucleus form the first shell—the K shell as it is called. Then comes the L shell, consisting of eight electrons ; next

FIG. 2

Models of the Helium and Neon Atoms

The nuclei are shown as large black circles and the electrons as small dots arranged in their respective shells. Each electron actually moves in its own orbit, as in Fig. 1

the M shell, with 18 electrons ; and so on into the N, O, P, and Q shells—the maximum number of electrons in any shell being $2n^2$, where $n = 1$ for K, 2 for L, 3 for M, etc.

Fig. 2 shows the atoms of helium (atomic number 2) and neon (atomic number 10) in which the K and the K and L shells, respectively, are filled ; the nuclei are represented simply by large dots. The figure is purely diagrammatic. Actually, of course, the dotted circles have no existence, and the electrons must be regarded as moving in orbits round the nucleus, and not as fixed at definite points. Although no two electrons occupy the same orbit, we may think of the K orbits as small and similar to one another, while the L

orbits, also similar to one another, are much larger. There are subdivisions in the L and larger shells which, however, we need not consider. Hydrogen, with only one electron, has that electron in the K shell, and the elements between helium and neon (viz. lithium, beryllium, boron, carbon, nitrogen, oxygen, and fluorine) have respectively 1, 2, 3, 4, 5, 6, and 7 electrons in the L shell, with a complete inner K shell of two electrons.

Chemical Families : Now it appears that elements belonging to the same chemical family have the same number of electrons in the shell which is incompletely filled. In the atoms illustrated in Fig. 2 there is no uncompleted shell, and these atoms belong to the family of rare gases. Lithium (atomic number 3) has a complete K shell and one electron in the L shell, and sodium (atomic number 11) has complete K and L shells and one electron in the M shell ; and we find that lithium and sodium belong to the family of alkali metals. The other alkalis—potassium, rubidium, cæsium—all have this characteristic of possessing one electron in a shell which can contain more. Similarly, the alkaline earths (beryllium, magnesium, calcium, strontium, barium) have two electrons in uncompleted shells.

Metals and Non-metals : Generally speaking, when the uncompleted shell is less than half full the elements show metallic properties, and when it is more than half full they show non-metallic basic properties. The rare gases, which have no uncompleted shells, show no chemical properties at all. The alkalis exemplify the metallic elements, while the halogens (fluorine, chlorine, etc.) have a shell lacking only one electron and are strongly non-metallic. The single electron of the alkalis is much more easily detached than one of the many electrons of the halogens, and the alkalis therefore tend to be positively charged and the halogens negatively. It is a useful rule to think of the atoms as possessed by a desire to have only complete shells of electrons round the nucleus. An element

with less than half the full complement would then find it easier to get rid of those than to fill the shell, while an element with more than half could more easily acquire the balance than give up those which it has.

The application of these general principles to particular cases involves, of course, much complexity, arising largely from the fact that we usually have to do with molecules of compounds instead of atoms of elements. We shall say something about the structure of molecules later, but meanwhile the picture of the atom drawn above serves to give us an insight into the kind of thing which occurs when we produce electrification by friction. It will be noticed that we have not *explained* the attraction and repulsion of electrified bodies. We have simply expressed it in terms of the attraction and repulsion of elementary particles. That, however, is sufficient for our needs, for, if we can describe what happens to observable and manageable bodies in terms which allow us to trace a connection between apparently dissimilar phenomena, we can do all that it lies within the power of science to achieve.

Magnetism

Further evidence of activity in matter is afforded by the existence of magnets. Everyone knows something of the property of certain pieces of iron or steel—and, to a smaller extent, cobalt and nickel—by which they can attract other pieces of iron and steel and hold them up against gravity; and there is, in fact, a naturally occurring oxide of iron, known as *lodestone*, which has the same property. If we suspend a magnet by its centre so that it hangs horizontally, and then bring the end of another magnet near one of its ends, we find that the suspended magnet is either attracted or repelled; while if we present the other end of the second magnet to the same end of the suspended one, the reverse happens—there is either repulsion or attraction. On the other hand, *either* end of the magnet will attract pieces of iron which are not magnets.

Magnetic Polarity

We have here a behaviour somewhat similar to that of electrified bodies. The magnitude of the force is again far greater than that of gravity; and there is the same attraction and repulsion between bodies affected, and only attraction between an affected and an unaffected body. We might therefore speak of positive and negative magnetization if we wish. We do not, however, use these terms, but speak of the two ends of the magnet as the *north* and *south poles*. The reason for this is that a freely suspended magnet always hangs so that one end points approximately towards the north and the other approximately towards the south, and if we disturb it, it always returns to this position. We therefore speak of the *north-seeking* and *south-seeking* poles, and these names are usually abbreviated to north (N) and south (S) poles. The rule is then found to be that unlike poles attract and like poles repel one another, while either pole "induces" in the nearest part of a piece of unmagnetized iron a pole unlike itself, and so brings about attraction. All this is quite analogous to positive and negative electrification.

There are differences, however. One is that, in suitable circumstances, any body can be electrified, but only a few kinds of body show the magnetic properties of iron. (We shall see that all bodies may show a much weaker and different kind of magnetization, but that can be ignored for the present.) Another difference is that an electrified body may be wholly positive or wholly negative, but a magnet always has a north and a south pole in the same piece of matter. It is clear, then, that we are not dealing here with a particular case of electrification, but with action of a different kind, which we must interpret in terms of our model of the atom if we are to use that model effectively.

Interpretation of Magnetism

The interpretation given is as follows. We assume that an electron moving in an orbit behaves as a small magnet. For

simplicity, suppose the orbit is a circle in the plane of this sheet of paper, and suppose the electron is revolving in a clockwise direction. Then the upper side of the paper is a S-pole and the lower side is a N-pole. If another similar orbit existed in a parallel plane just above the first there would therefore be attraction between them and the orbits would approach one another, while if the second electron were revolving in the opposite direction to the first the polarity would be reversed and there would be repulsion. Each atomic electron revolving in its orbit is therefore a small magnet, and the magnetic properties of observable bodies must be expressed in terms of the interaction of these intra-atomic magnets.

Like the assumption of the existence of elementary particles in atoms, this is not arbitrary guesswork. We can experiment with electrically charged bodies of observable size moving in orbits, and we find that they do in fact behave as magnets in the manner just described. It is therefore quite reasonable to suppose that the elementary charges behave similarly, and provide us with the elements out of which we can build a satisfactory theory of magnetism.

Magnetic and Non-magnetic Substances

It may not at once be obvious why a few elements should show strong magnetic properties, while the remainder show scarcely any. The reason in detail is beyond our scope in this book, but since, in general, an atom contains many revolving electrons, and their orbits may have various orientations in space, it is not difficult to understand that the force due to one " orbital magnet " may be weakened or neutralized by that due to another acting in the opposite direction, so that the net magnetic effect of all the electrons in the atom may be large or small, depending on the way in which the orbits happen to be arranged. Further, even if each atom has a fairly strong resultant magnetism, the directions of the forces due to the many atoms in a piece of matter of visible

size will in general be so varied that there will be little, if any, resultant magnetism to be observed. Special conditions within the atoms are necessary for magnetizability, and special arrangements of the atoms with respect to one another in a magnetizable substance are necessary for it to behave as a magnet. This explains why only a few elements are magnetizable, and why all samples of those elements are not magnets.

Light

A third piece of evidence that matter can be active comes from the fact that we can see it. In ordinary daylight we see our surroundings, not through any original activity on their part but because of the Sun. But the Sun is a piece of matter, which could presumably be as dark and inert as the Earth in certain circumstances. The fact that it is not so we attribute to its power of emitting something which we call "light," which enters our eyes and thus enables us to see it. The light falls also on other bodies, and is then reflected to our eyes, so that we can see them also. That such a thing can happen in ordinary terrestrial matter as well as in the Sun we know from our experience with the filament of an electric lamp, for instance. When the current is switched off the filament is inactive, and in the absence of light from other sources we cannot see it. Switch on the current, and it immediately becomes luminous and illuminates other bodies. The filament can therefore be either passive or active, according to circumstances, and again we express its activity by saying that it emits light.

We shall not at this stage enter into the question of what light is. We may, however, find ample justification for our assumption that there is actually some objective thing discharged from a luminous body in the fact that objective events occur in the neighbourhood of such bodies, such as chemical action in a photographic plate. What we have to

do now is to see if our picture of the atom enables us to express the emission of light in suitable terms.

Process of Emission of Light

The clue to the explanation lies in the fact that in order to make a dark body luminous we must always supply energy. Heat energy is usually the most convenient form, and will serve as a sufficient example for our purpose. We know, for instance, that when we heat a poker by putting it in the fire it becomes luminous and gives out a red light. In accordance with the principle of conservation of energy, therefore, we are led to regard light as a form of energy into which heat energy may be transformed. We look to the atoms of the body which becomes luminous for an understanding of the mechanism by which the transformation is brought about, and we find that a satisfactory explanation can be found.

Heat energy, as we know, is the kinetic energy of the atoms of the hot body. As the temperature rises the atoms collide more and more vigorously, and so impart their energy to one another. Now if the atoms were simple particles this would mean that their velocities would go on increasing and so the body would go on getting hotter. But actually it appears that the energy imparted to an atom by a collision does not always move the atom as a whole, but may be used in displacing an electron into another position; the electron, instead of revolving in its normal orbit, may be pushed out to an orbit farther from the nucleus and revolve there instead. In that position, of course, it has greater potential energy than it had before, just as a stone has greater potential energy if it is raised above the surface of the Earth. The first stage in the transformation, then, is the conversion of the heat energy of the colliding atom into potential energy of an electron in the atom which is struck.

But the new state of the latter atom is not a stable one. Just as a stone thrown into the air returns again to the Earth,

so an electron removed to a more distant orbit returns as soon as possible to its original place. This is another example of the tendency of a mechanical system to come to the position of minimum energy (see I, 67).* It has been calculated that the time the electron remains in the larger orbit is only about one hundred-millionth of a second. On returning, of course, it loses the potential energy it had gained, and the question is, what becomes of that energy? The answer is: it is radiated as light. How such a transformation can occur we cannot visualize in detail. It is sufficient for our purpose to know that it does occur, and when it happens the body in which it happens becomes luminous. Many atoms, of course, act in the same way at the same time, so the light we observe is made up of the radiation from large numbers of them.

Atoms as Energy Transformers

This power of transforming energy of other kinds into light is a fundamental property of the atom. There is no other way in which light can be produced; whenever anything is visible to us we know that atoms are at work transforming energy of some kind into the light by which we see it. The transformation can be reversed, as we know by the feeling of warmth we get when standing "in the Sun." The light energy falling on us is transformed back again into atomic (or molecular) motions in our bodies, and this constitutes heat.

Photons and Ether Waves

Although we are not now concerned with the nature of light, it will be convenient to mention here that there are two views of the matter which, though they appear to be irreconcilable with one another, are, each in its own way, exceedingly useful in giving us a vivid representation of the phenomena of light and enabling us to grasp complex facts

* References to the first volume of this course, bearing the title *Mechanical Physics*, are indicated by the symbol I, followed by the page number.

in a simple, comprehensive way. According to one, the light emitted from an atom may be regarded as a small particle (it is usually called a *photon*), something like an electron but with no charge, extremely small mass, and perpetual motion at a very high speed. When we are dealing with the effect of light on the individual atoms of matter which it encounters, and its transformation back into other forms of energy, this is the view usually adopted. The other view is generally taken when we are concerned with the passage of light through empty space or through transparent bodies ; light on this view consists of transverse waves in an all-pervading medium called the *ether* (see I, 103), and these waves spread out in all directions from the luminous body with the highest velocity known in nature (3×10^{10} cm., or 186,000 miles, a second). The frequency of the waves is very high, and the wave-length therefore extremely small, ranging from about 4 to 8 hundred-thousandths of a centimetre.

We shall not attempt to discuss which, if either, of these views is the "right" one. Our aim, as we have said before, is first of all to understand what happens among observable things, and to find connections between perceptible occurrences which seem at first to be independent of one another. What happens in the unobservable world is of interest to us only in so far as it serves that purpose, and we shall accordingly speak in terms of either photons or light-waves, according to the need of the moment. We have at least this safeguard, that the two views never lead to contradictory results. When we explain something in terms of photons, the wave theory does not give an alternative explanation—it gives none at all ; and *vice versa*.

RADIO-ACTIVITY

There is one other activity of matter which must be mentioned here—namely, that known as *radio-activity*. Certain elements—mainly those of highest atomic number—are not

the stable things that we are accustomed to think elements to be, but are continuously and spontaneously radiating something into space. We know this because, among other things, photographic plates in the neighbourhood of the bodies are affected, and we cannot attribute this to light because, although the elements in question are sometimes visible in the dark, the plates are affected even when protected from light. On examination the radiation turns out to be of three different kinds : there are first of all nuclei of helium atoms, next electrons, and thirdly radiations similar to light but of very much higher frequency, and therefore smaller wave-length. These three kinds of radiation are known respectively as α, β, and γ rays.

Transmutation of Elements

The best known and most active of radio-active elements is radium, but there are several others. We can give only the barest account of the phenomena here. The elements which show them form series, which are generated successively as higher members break down into lower ones. Radium, for example, spontaneously changes into radon, and this again into a succession of other elements, ending in the stable element, lead. This transmutation of elements, resulting from the loss of the particles radiated, is a process which is quite unaffected by any changes of temperature or pressure or anything else that we can apply in the laboratory, though it is believed that under the extreme conditions existing in the interiors of stars the process is somewhat modified.

Explanation of Radio-activity

Our model of the atom accounts for radio-activity in terms of the spontaneous breaking up of atomic nuclei. Clearly some such explanation is necessary, for the α-rays—or *α-particles*, as they are often called—consist of helium nuclei, which are each made up of two protons and two neutrons, and these particles normally exist nowhere but in the nuclei

of atoms. When an atom of radium, for example, discharges an α-particle, its nuclear charge is reduced by two units and its mass by four units. It accordingly becomes the nucleus of an atom of lower atomic number (the element radon, the atomic weight and atomic number of which are respectively 222·4 and 86, while those of radium are 226·4 and 88), and two revolving electrons, now become superfluous, leave the atom. The β-rays are simple electrons, but the bulk of them are not, as might have been expected, the discarded orbital electrons just referred to. Their origin is obscure, but it appears that they are created in the nucleus, by some process not yet fully understood, out of the protons and neutrons there. It is unnecessary for our purpose to enquire further into this, which forms the subject of much experimental research nowadays. The γ-rays are similar in nature to light rays, and travel with the same speed, though they do not stimulate the sense of sight. They, like the electrons, are generated during the changes which take place in the atomic nuclei. We think of them, as we do of light, as being either particles or waves, according to the problem under consideration ; but as particles they have much greater energy, and as waves much higher frequency, than light. This appears to be the only difference between γ-rays and light, but it makes a great difference to the physical effects of the two kinds of radiation.

Radiation in General

The term "radiation" is used in a general sense to indicate anything not directly perceptible that travels through space. We have to say "not directly perceptible" because all bodies travel through space, and we do not usually refer to stars and planets and ordinary pieces of matter as radiation ; nor, in fact, is it customary so to refer to sound or other waves in matter, though literally these would come under the definition. The term is restricted to light and such things

as γ-rays which appear to be identical with light in everything but frequency (speaking in terms of the wave theory), on one hand, and to elementary particles such as electrons and protons not formed into atoms on the other. Small collections of particles, such as α-particles, are also called radiation, or rays, if they do not make complete atoms.

Cosmic Rays

It may be true to say that all radiation is emitted from single atoms (or molecules) of matter, but we cannot be sure of this, since we find that everywhere accessible to us there is a small amount of a very penetrating radiation known as *cosmic rays*, the origin of which is unknown. As the name implies, they are believed to exist throughout the universe. It is difficult to be sure of their fundamental nature because, when they enter the Earth, they are apt to release particles from atoms (*secondary* rays), which become associated with them and are hard to distinguish from the primary cosmic rays. They include, as well as radiation of some of the kinds already mentioned, particles of kinds not so far referred to (*positrons*, *mesons*, etc.), which we shall not need to describe.

The Velocity of Light

There is one fundamental distinction between the various kinds of radiation—namely, that certain types, when travelling through empty space, do so with a definite constant speed—the speed of light. This speed, as we have said, is the highest found in nature, and is usually denoted by the symbol c. The other types of radiation always move with a lower speed than this and their speed has no definite value—it depends entirely on circumstances. They can be accelerated or retarded, and presumably could be at rest, though we should then have no means by which to observe them. The first type of radiation includes light- and γ-rays, as well as various other kinds, which we shall mention later. It is this kind of radiation that can be interpreted as ethereal waves, for

the velocity of a wave is determined by the medium in which its undulations occur and not by the circumstances of its emission (I, 195). The other kind (including protons, electrons, α-particles, and the non-ethereal constituents of cosmic rays) are usually described as particles, but in certain phenomena these too appear to behave as waves, though not waves in ether. This view of them, however, is necessary only for certain special investigations which are outside our scope, and we shall not need to think of them as other than particles.

Spectrum of Ether Waves

The ethereal wave radiation is classified according to its frequency or wave-length. Fig. 3 shows, diagrammatically, the various divisions which have been made, based on the physical effects of the radiation. Waves whose length (λ) falls between 4×10^{-5} and 8×10^{-5} cm. (and whose frequency, n, therefore lies between $\frac{3}{4}\times 10^{15}$ and $\frac{3}{8}\times 10^{15}$, since $c = n\lambda$ (I, 195) and $c = 3 \times 10^{10}$ cm./sec.) form visible light. Within this range, difference in λ or n is visible to us as difference of colour. The shortest waves are violet and the longest red, and the intermediate colours are in the sequence familiar in the rainbow, giving violet, indigo, blue, green, yellow, orange, red as the order of colours in the ***spectrum***, as it is called. The limits for visibility are rather indefinite, like the limits for audibility of air-waves (I, 222), and vary for different observers, but the values given here are not far from the average for normal persons.

Waves shorter than violet light are known as ***ultra-violet***. They affect photographic plates and cause many bodies to emit " fluorescent " light (see p. 119). Those as short as 10^{-8} cm. or thereabouts are known as *X-rays*, and the γ-rays already mentioned are still shorter. At the other end of the spectrum, waves longer than red light are known as ***infra-red***, and when they reach lengths of a few metres they become

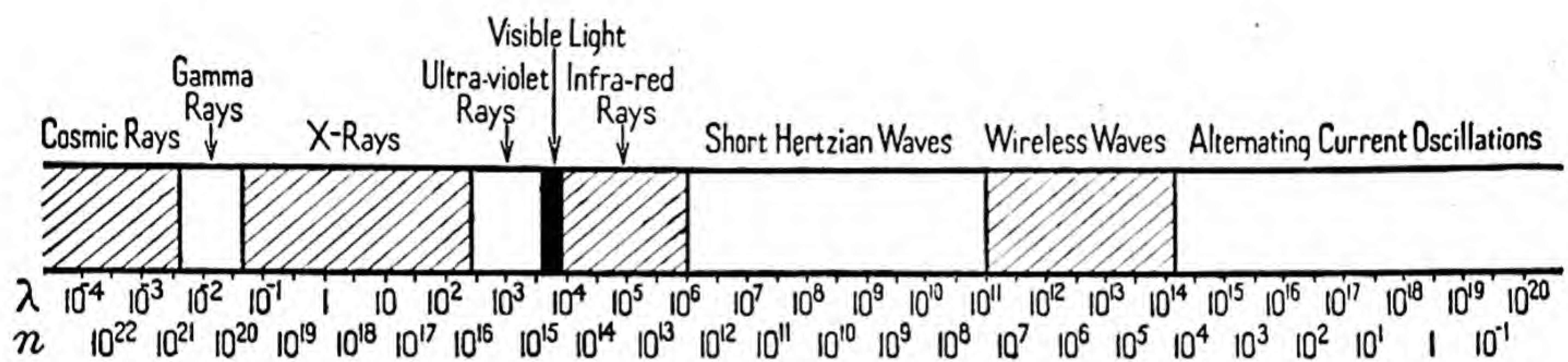


FIG. 3

Analysis of Ethereal Vibrations

The scale shows the wave-length (λ) in angstroms (1 angstrom = 10^{-8} cm.) and frequency (n) in vibrations per second, plotted logarithmically, *i.e.* so that equal divisions correspond to equal multiples of λ or n. The divisions between waves of different types are not abrupt, and the boundary lines merely indicate approximately where waves of one type merge into those of another

effective in wireless telegraphy, and are familiar to everyone as *wireless waves*.

Properties of Ether Waves

It will be clear that, although the difference between these various ethereal waves is simply one of wave-length or frequency, this shows itself in profound differences of physical properties. One very important difference is that of penetrability into matter. As everyone knows, some bodies are opaque to light and some transparent, while others (*e.g.* ordinary dark-room red glass) are opaque to some colours and transparent to others. In the near ultra-violet, bodies tend to be on the whole more opaque ; thus, glass will not transmit far below the violet limit, and when we get down to wave-lengths just below 2×10^{-5} cm. very few bodies allow the radiations to pass—even air becomes opaque. In the X-ray and γ-ray regions, however, there is a great change, and these very short waves will pass through at least small thicknesses of almost every kind of matter. In the near infra-red, again, we get an increase of opacity, and when we come to wireless waves we find very curious phenomena—glass being opaque, and such things as pitch and soot transparent.

Velocity of Ether Waves in Matter

It has been said that the velocity of all kinds of ethereal waves is the same in empty space. That is no longer true when the waves enter a transparent body. The velocity then is, in general, reduced, and the amount of reduction depends on the wave-length. There is no universal rule, but with visible light, which enters most commonly into our experience, the greater the wave-length the less in general is the reduction of velocity. Thus red light is retarded less than violet light on passing through glass or water. This, however, in no way affects the sharpness of the distinction between radiation which can be called ethereal and that usually described as particles, for the velocity of the former is no more under our

control when the waves travel through matter than when they travel through empty space. A given kind of ethereal radiation will travel through a particular kind of matter in a given physical condition at its own unalterable speed, whereas the speed of the particles can be varied at will.

Planck's Constant

It is probably true to say that all ethereal radiation is emitted from atoms or molecules in the manner already described, by the absorption of energy in some form and its transformation into radiation. The frequency, n, of the radiation emitted in an atomic transition depends on the amount of energy, E, which it contains in a very simple and remarkable way, represented by the equation

$$E = hn \qquad . \quad . \quad . \quad . \quad . \qquad (1{\cdot}1)$$

where h is a very small constant quantity known as *Planck's Constant*, having the value $6{\cdot}55 \times 10^{-27}$ erg-seconds. That is to say, if an electron removed from its normal position in the atom to a larger orbit falls back to a lower orbit where it can do with E units of energy less, those E units will be sent out into space as a wave (a *quantum*, it is called) of ethereal radiation of frequency $n = \frac{E}{h}$. If E is very large, then n will be very large (*i.e.* λ will be very small), and we shall have an X-ray or a γ-ray. If, on the other hand, E is very small, we shall have an infra-red wave or a wireless wave. The wireless waves, in fact, can be, and usually are, produced by the movements of ordinary electrically charged bodies, though ultimately, of course, the atoms supply the charges ; but to obtain the shorter waves we have to supply matter with energy in some other form and let the atoms make the transformation.

General Remarks on Atomic Theory

The purpose of this chapter has been to give a general account of the sub-atomic structure of matter, which the reader can

keep in mind when trying to understand the various facts to be described later. It throws much light on them, but it should be said that the details of electrical, magnetic, and optical phenomena do not admit of so ready an explanation in terms of the structure of the atom as do mechanical phenomena in terms of the simple molecular theory. This is partly because sub-atomic theory is as yet far from complete—and, indeed, as we have seen, is in some respects unsatisfactory in that it demands different, and apparently irreconcilable, views of the nature of the ultimate constituents of matter. But it is also in part due to the fact that the explanation, when it can be given, is sometimes so intricate that for practical purposes it is more expeditious to learn the facts in terms of practical rule-of-thumb statements than as consequences of the fundamental nature of matter.

The picture of the atom and its behaviour which we have given is the one which enables us to explain the largest number of facts without becoming too abstruse. The latest theories explain more, but are much less easily grasped, present no imaginable picture of the atom, and are, moreover, liable to supersession by still more comprehensive theories as progress continues. The reader must understand that we do not say the atom is exactly as we have pictured it; it appears, in fact, that a precise picture is impossible to give. The value of the view we have adopted is that it gives a satisfactory and not misleading account of the facts which concern us in practical work. Where it fails we shall simply describe the facts without attempting to explain them, or adopt some temporary model of the underlying causes which will be valid so far as we need to apply it.

There is a considerable part of our subject which can be described without specifying more than the most general characteristics of the imperceptible agencies responsible for what we perceive. For example, we can go a considerable way in the subject of optics without requiring to know anything about the nature of light except that it is something

which travels outwards from the luminous body. Whether it consists of waves or particles or something unimaginable makes no difference so long as we are concerned only with the path which it takes in specified circumstances. Similarly, we can describe many of the phenomena of magnetism without concerning ourselves with the way in which the electrons in atoms contribute to the resultant magnetic force : it is sufficient to know that an atom behaves as a small magnet. We shall make use of the simplification which this allows. In short, our excursion into the interior of the atom is to be regarded as a purely practical adventure, undertaken for the sake of the facility which it affords us of understanding what is happening in the observable world, and not as a search for ultimate truth which we must introduce into the explanation of everything we encounter. When it helps us we shall use it ; when it does not, we shall ignore it.

We begin with the subject of optics, and, except for the last chapter, this section will be simply a geometrical treatment of the path of light in optical systems. We next consider electricity, at rest and in motion, and the magnetism of permanent magnets, at first without reference to any connection between electricity and magnetism. Later, electro-magnetism and the terrestrial aspects of electricity and magnetism are discussed.

EXERCISES

1. Give a general account of the structure of the atom, and explain how the atom of one element differs from that of another. If a neutron were added to the nucleus of an atom, how would you expect the chemical properties of the atom to be affected ?

2. Explain the production of electrification by friction in terms of atomic structure. Why is it that, when two bodies are charged by being rubbed together, they never both acquire the same kind of electrification ?

3. Why can magnets be made from some elements and not others ? How is magnetism associated with the structure of matter ?

4. When a substance is made very hot it becomes luminous : explain why this occurs. If two atoms radiate, by single transitions of electrons, beams of red and blue light respectively, which beam will contain the greater amount of energy ?

5. Describe the phenomena of radio-activity, and show how they are interpreted in terms of the structure of the atom.

6. Give a general description of the kinds of radiation that have been observed to proceed from matter in various circumstances, and explain how those which can be attributed to waves in the ether are distinguished from one another.

PART II

OPTICS

CHAPTER II

EMISSION AND REFLECTION OF LIGHT

Rectilinear Propagation of Light

IN this section we make the assumption that our sensation of sight is due to the emission from luminous bodies of something called *light*; and we see what we can learn from simple experiments about the way in which light behaves when it meets matter, without concerning ourselves with what it is. If it helps us to think of light as a lot of small particles travelling outwards from the luminous body in all directions, we may do so; it will make no difference to the facts we have to describe.

The first thing we notice about light is that it travels in straight lines. The simplest evidence of this is the fact that if we want to see a body we must look straight towards it. If we lay a straight rod between the body and our eye, we must look along the rod or we do not see the body. The formation of shadows gives further evidence. In Fig. 4, ABCD is an opaque square, S a point-source of light, and LMNP a white screen. The shadow of ABCD—namely, A′B′C′D′—has the outline given by drawing straight lines from S to the edges of ABCD, and producing them to the screen. There is no light in A′B′C′D′, because the light which passes outside the square travels in straight lines, and so cannot get there.

Careful observation shows that, to a very slight extent,

light does creep just inside the shadow. This effect, known as *diffraction*, we shall mention again later ; at this stage we shall ignore it. It must not be confused with the blurred edges of the shadow which we get when the source at S has a finite size. That is caused by the fact that the source consists effectively of many points S in different places, and so

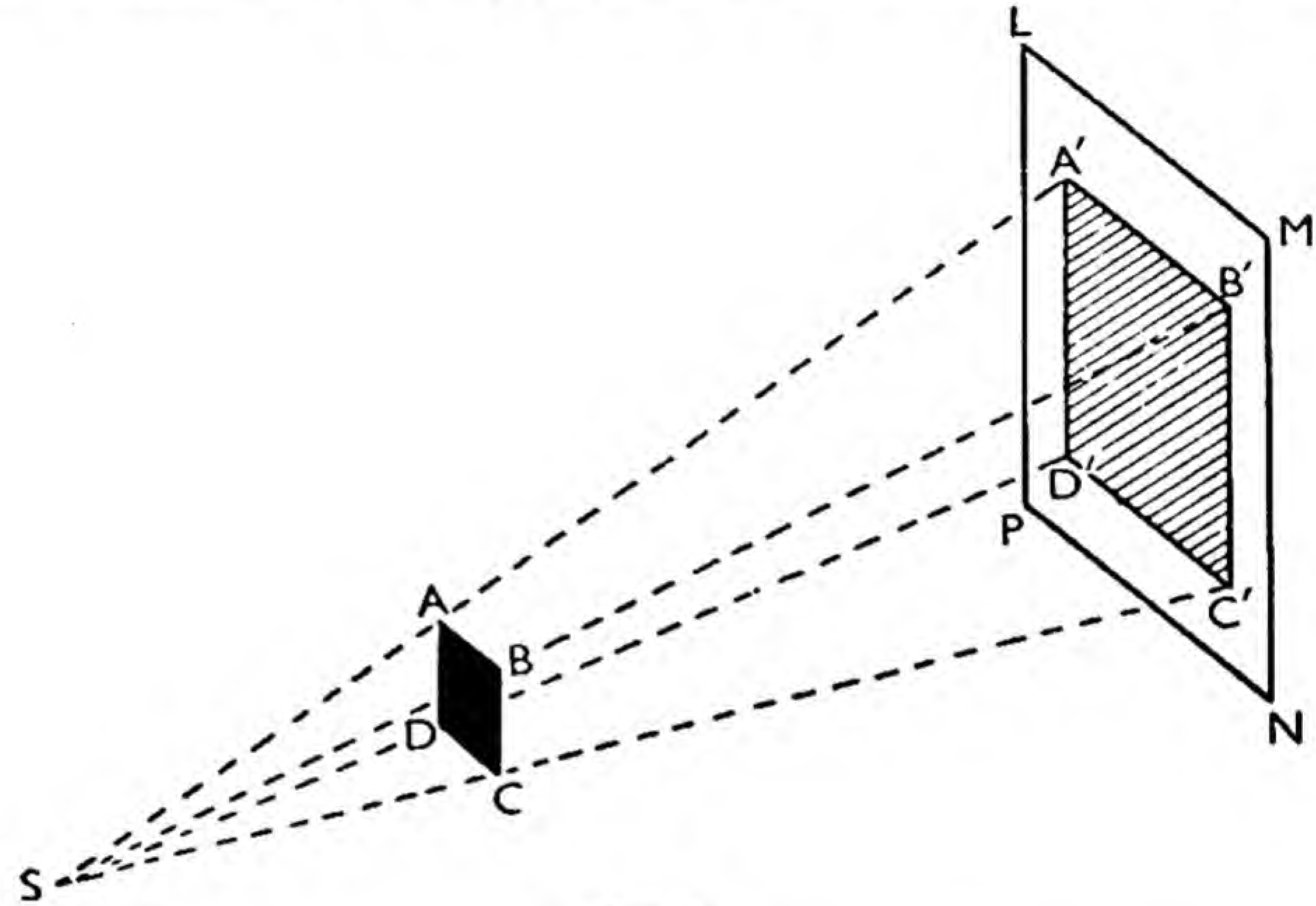


FIG. 4

Shadow (A′B′C′D′) of opaque object (ABCD) on Screen

there are a large number of shadows which overlap one another.

We can often see this *rectilinear propagation* of light, as it is called, when sunlight shines through the window of a room, and we seem to see the rays passing through the air in straight oblique lines. What we actually see is not light but particles of dust in the air which are illuminated by the light. This we can prove by letting the light pass through a glass box containing dusty air, when the path inside will be visible, and then exhausting the box by a pump. The light still passes through, and may be seen emerging at the far end, but its

path in the box disappears. Light itself is quite invisible: we see matter by means of light, but we do not see light.

We may picture a luminous body, then, as the origin of straight lines issuing from it in all directions. Any one of such lines is called a *ray*, and a small bundle of them is called a *pencil* or *beam* of light. Strictly speaking, there is no such thing as a ray of light, because light, whatever it may be, must have some thickness, and a geometrical straight line has none. A pencil of light is the actual physical thing that exists. Nevertheless, it is often convenient to speak of rays of light, but we shall see instances in which this practice will lead us into error unless we are careful.

Light and Matter

When light falls on a piece of matter, any of three things may happen to it. First, it may be *absorbed*, *i.e.* transformed into some other form of energy, such as heat, in the body. Secondly, it may be *reflected* or *scattered*, *i.e.* sent out again into space, generally in a different direction. Thirdly, it may be *transmitted*, *i.e.* pass through the body with or without a change in direction. Usually, more than one of these things occur, part of the light behaving in one way and part in another. Thus, if we look into a shop window we may see the Sun reflected there, and we know that if we were inside the shop we should see sunlight passing through the window. If we were to add together the light reflected and that transmitted, we should find that the sum was less than the amount of light falling on the window, so that some must be absorbed in the glass—and, in fact, the glass is to some extent warmed by the sunlight which it absorbs.

We shall not concern ourselves with the absorption of light, because when it is absorbed it is no longer light. We have said something on this subject in connection with the absorption of radiant heat (I, 185), and what was said there applies to light also. We must, however, consider the scattering, reflection and transmission of light.

Scattering of Light

Suppose we are in a shuttered room, in which we can switch on an electric light at will, and we look towards some object—say, a sheet of paper. When the light is switched off we see nothing, so the paper does not emit light. When the light is switched on, however, we see the paper, and we see it from any point in the room at which we choose to stand. This means, first, that the light by which we see it must come originally from the electric lamp, fall on the paper, and then travel to our eyes; and secondly, that it must leave the paper

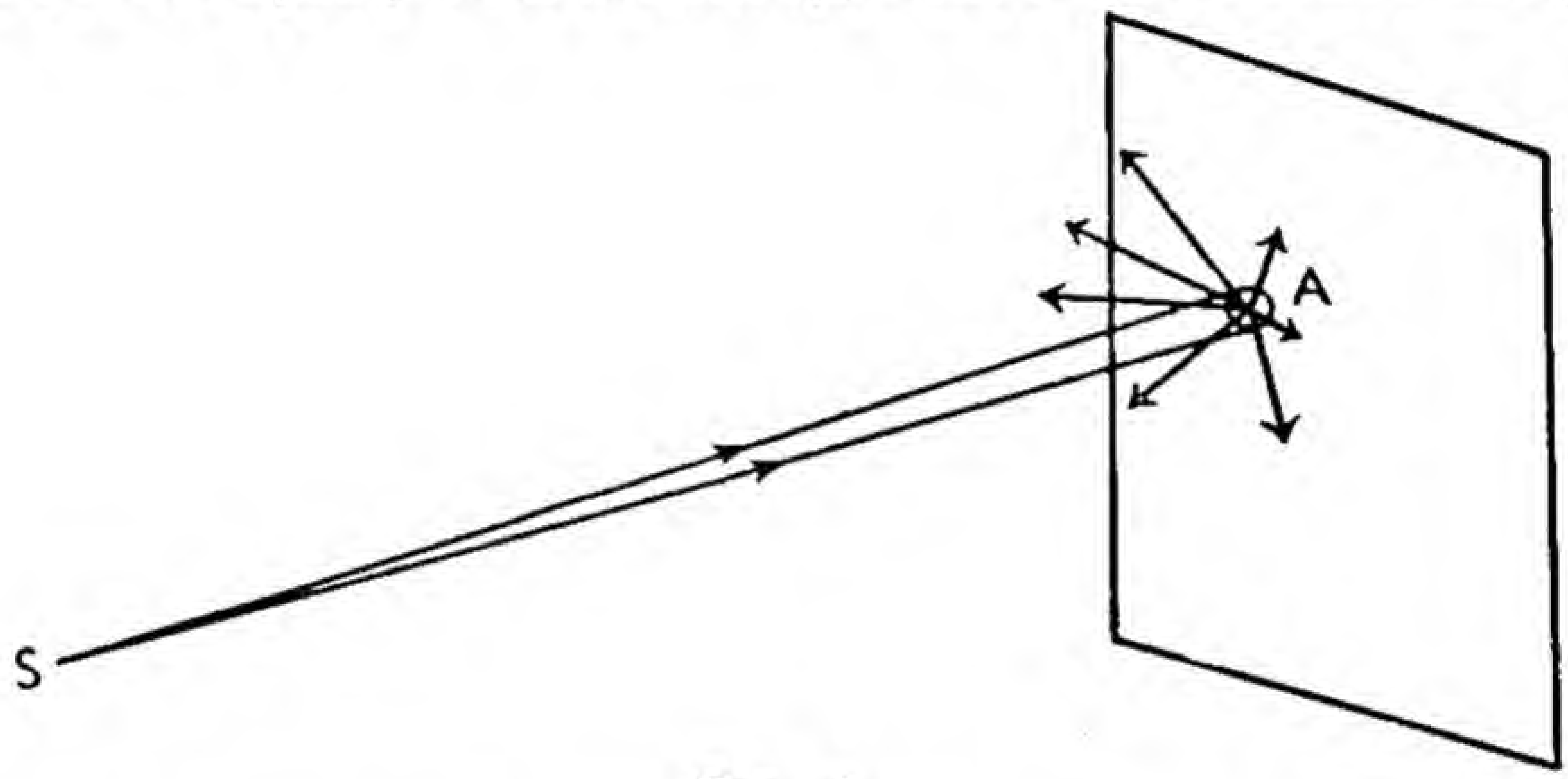


Fig. 5

Pencil of Light from Source S scattered from element A of a diffusely reflecting Screen

in all directions. The light is said to be *scattered* by the paper, and this is the process by which we see all ordinary objects; they scatter in all directions light which falls on them from some luminous body, and some of it enters our eyes as we look towards them. Fig. 5 illustrates the process for our sheet of paper. Each small element of the paper, A, is illuminated by a narrow pencil of light from the lamp, S, which then travels outwards, always in straight lines, in various directions from A, so that some of it meets our eye from whatever point we look towards A.

We may remark here that all bodies scatter light to some extent. Even air on the clearest day scatters a great deal of

the sunlight which passes through it, and it is this which gives us the appearance of a bright sky. If there were no scattering we should see a still brighter Sun in a sky otherwise black except for the stars and other heavenly bodies. It is the intensity of scattered sunlight that " puts out " the stars in the daytime, and the sky appears blue because some of the colours in sunlight are scattered more than others.

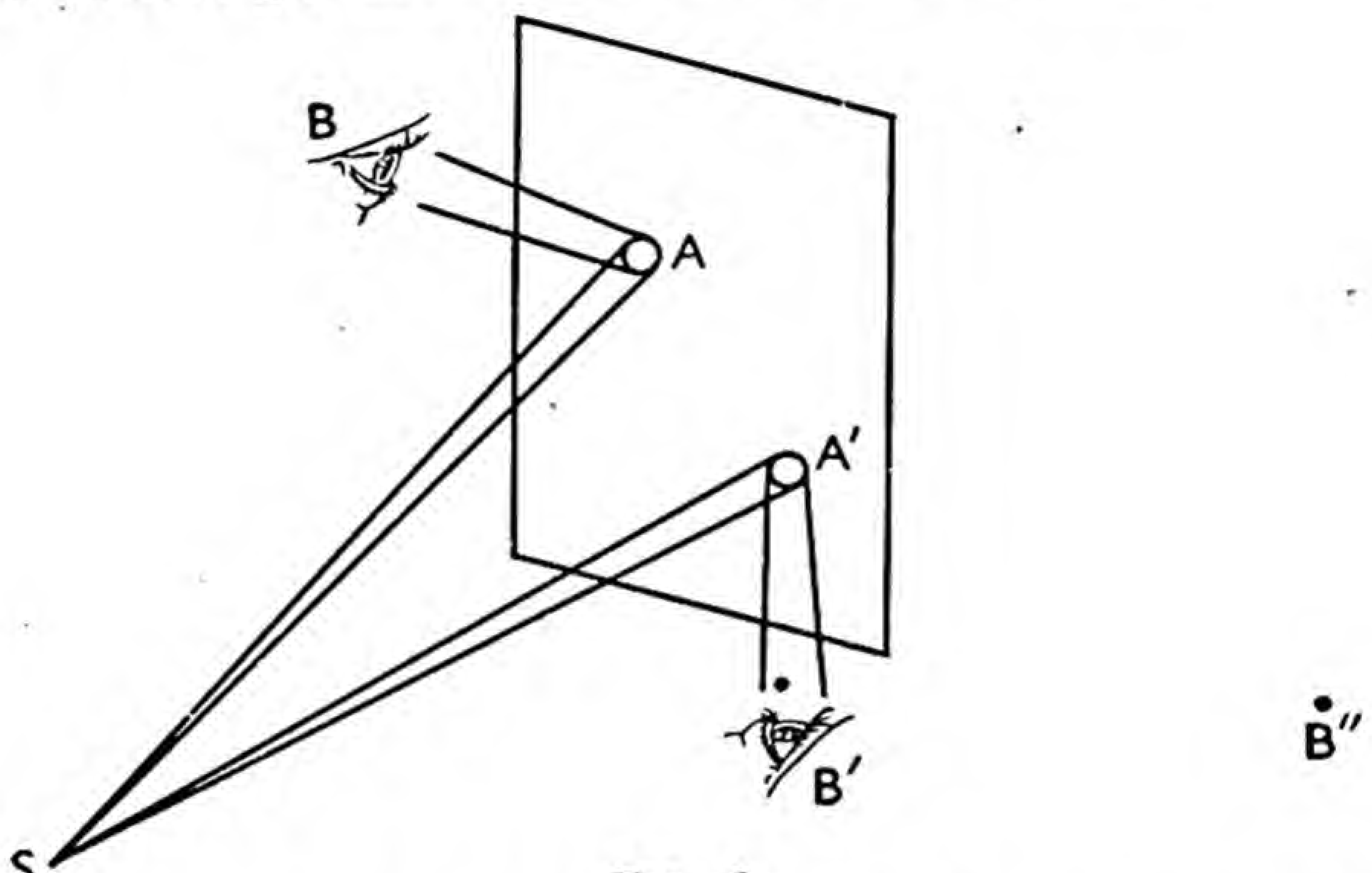


FIG. 6

Pencils of Light from Source S reflected from elements A and A′ of a Plane Mirror

Regular Reflection of Light

Returning to our room, suppose we substitute for the sheet of paper a very clean mirror of the same shape. We now see, not the whole, evenly illuminated square, but simply an image of the lamp, apparently situated at one place on or behind the mirror. As we move, the position of the image seems to move also, and from certain positions we do not see the image at all.

We can explain this by supposing that the light falling on each element of the mirror is not scattered in all directions, but reflected only in a certain definite direction. For example, suppose that when the eye is at B (Fig. 6) and looks towards

the mirror it sees an image of the lamp in the direction of A. Then the light falling on A would be reflected only towards B. If the eye is moved to B′ it sees nothing at A, but sees an image now in the direction of A′. The light falling on A′ must then be reflected towards B′. If the eye is at B″ it sees no image, because none of the light falling on the mirror is reflected towards B″.

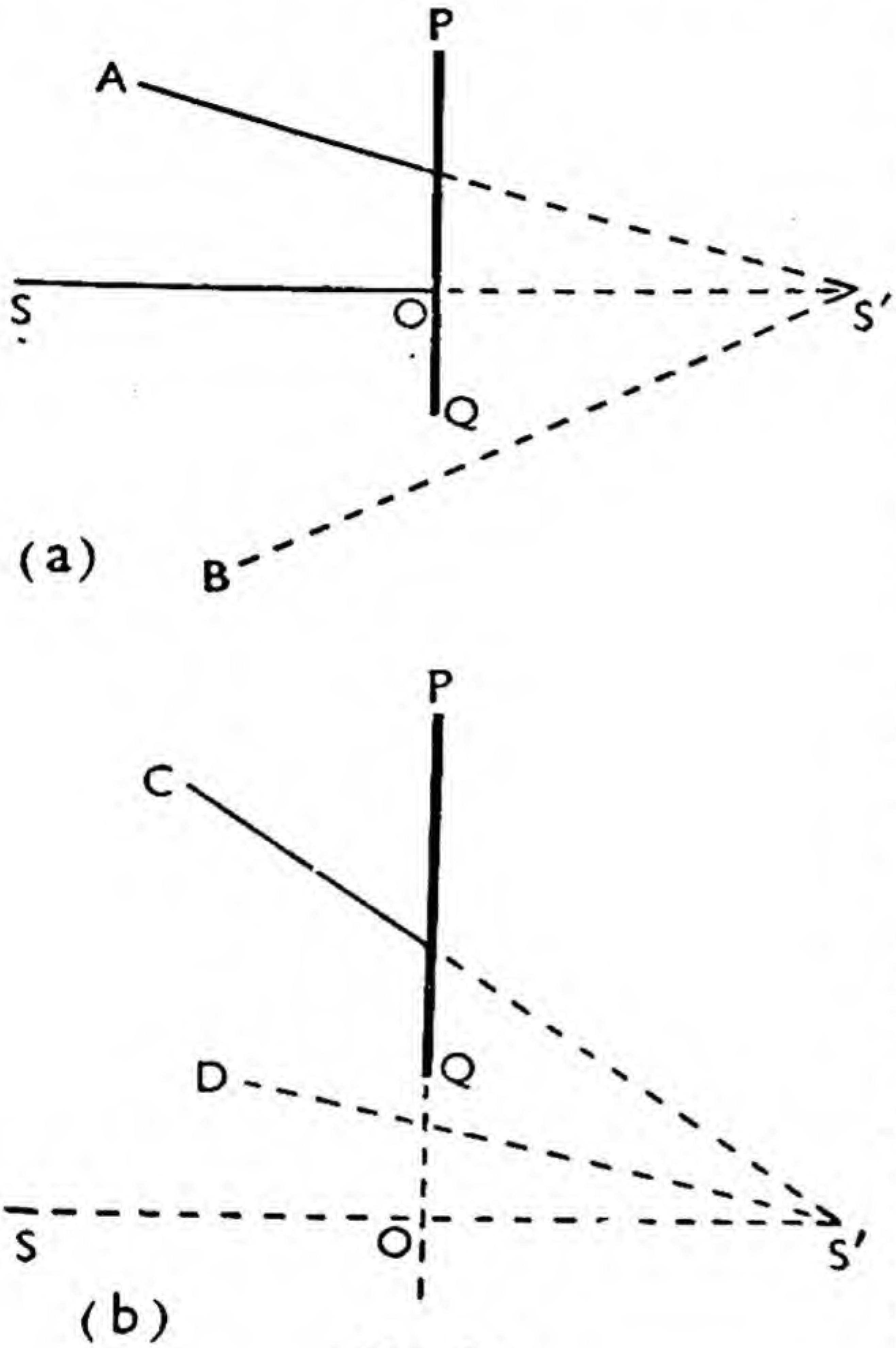


Fig. 7

Visibility of Image S′ of Source S in Mirror PQ

(*a*) Image visible from A but not from B
(*b*) Image visible from C but not from D

Law of Reflection

Now there must evidently be some rule by which we can determine whether we can see an image of the lamp from a given point, and, if so, in what direction we must look in order to see it. The rule, deduced from experiment, turns out to be this. Suppose (Fig. 7 (*a*)) PQ represents a section of the mirror and S represents the lamp. Draw a perpendicular SO from S to PQ, and produce it to S′, making S′O = SO. Then, in order to see the image we must look towards S′, and we shall see it if the line of vision cuts the mirror, but

not otherwise. If the eye is at A, for example, the image will be seen in the direction AS′, for the line AS′ passes through the mirror. If, on the other hand, the eye is at B, the image will not be seen at all, because the line BS′ passes outside the mirror.

If the line SS′ does not cut the mirror, the rule is not altered. We then draw SO perpendicular to PQ produced (Fig. 7(*b*)) and proceed as before. The same rule tells us that we shall see the image from C but not from D.

We say that the image of the lamp is situated at S′, because that is the point towards which we must look in order to see it. The rays of light which enter our eye proceed as though they came from S′. They do not, of course, but the law of reflection is such that they always behave as though they did. Now we can prove that, for this to be so, the law of reflection must be the same as that for sound (I, 205), viz. that *each ray incident on the mirror is reflected in the plane containing itself and the normal to the surface at the point of incidence, and the incident and reflected rays make equal angles on opposite sides of the normal.* These angles are called the *angle of incidence* and the *angle of reflection*, respectively.

The proof of this rule is as follows. Let M (Fig. 8) be the point in which AS′ cuts the mirror. Then we have to prove that the angles SMN and AMN (where MN is perpendicular to PQ) are equal. Now, since SOS′ is perpendicular to MO, and SO = S′O, the angles MSO and MS′O must be equal. But, since NM is parallel to SS′, MSO = SMN and MS′O = AMN. Hence SMN = AMN, and these are the angles of incidence and reflection, respectively. That the lines SM, NM, and AM lie in the same plane is obvious, for we take the plane of the paper to be the plane through S, S′, and A, whatever it may be, and the normal MN must then clearly lie in the plane of the paper also.

The actual path of the ray of light considered is then from S to M and from M to A. The lines OS′ and S′M are merely parts of the geometrical construction ; no light travels

along them. Furthermore, we have chosen the point A at any arbitrary position, so that the same law of reflection applies

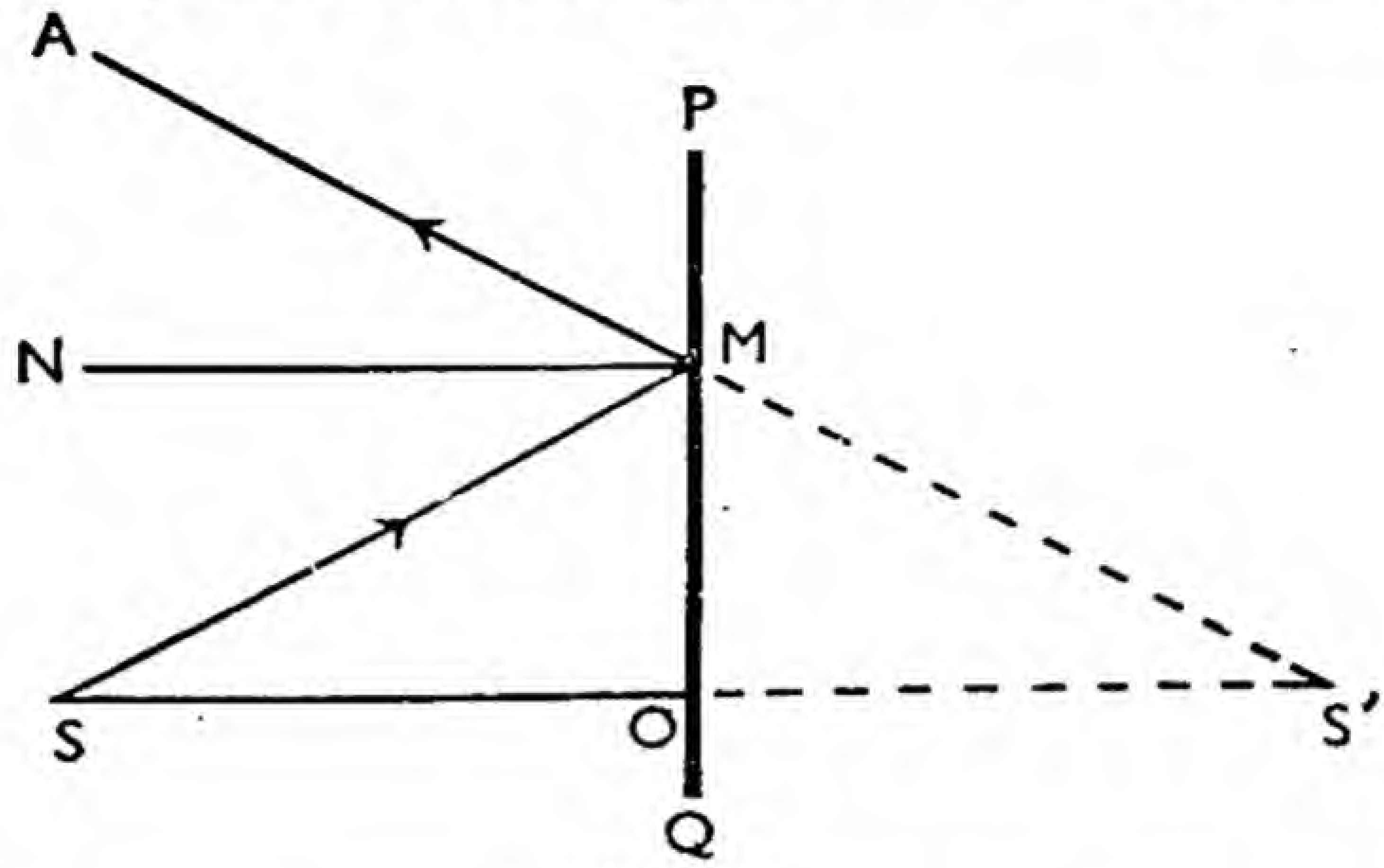


FIG. 8
Reflection of Ray SM along MA by Mirror PQ

to all rays proceeding from the lamp to the mirror. As we have remarked, what actually exists in nature is not a single

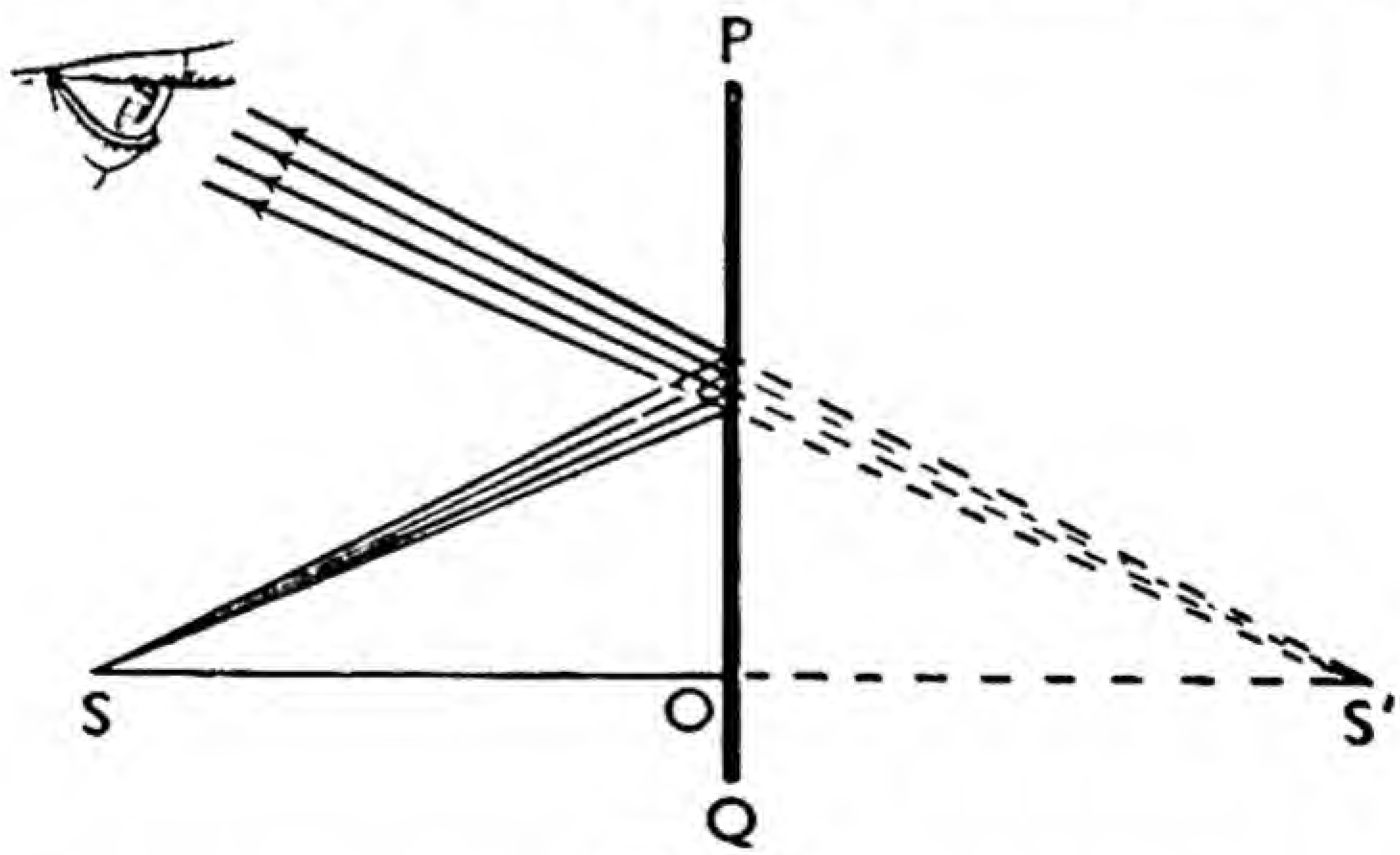


FIG. 9
Reflection of Pencil of Rays from Source S by Mirror PQ, giving appearance of Virtual Image at S′

ray but a pencil of rays—in this case a pencil large enough to fill the pupil of the eye. The path of such a pencil is shown in Fig. 9. Every ray in the pencil is reflected according to

the same law, and all proceed to the eye as though they came from S′.

Images

We thus have the following definition. The point from which the rays of light from an object appear to come after reflection from a mirror is called the *image* of the object. What the eye sees, of course, is what it would see if the unobscured object itself were at the position of the image; and if the mirror were perfectly clean and plane, and we could not see the object by direct view, we should not know from sight alone that the object was not actually at S′, for everything appears as though it were. The illusion breaks down, of course, when the line of sight passes outside the mirror, for then there are no reflected rays to cause it. We may here note one important thing, however, and that is that however many reflections and changes of direction rays of light from an object may undergo before entering the eye, the object always appears to be in the direction in which the rays finally enter the eye.

The image of the lamp at S′ is called a *virtual* image, because the rays only appear to diverge from it, and there is actually no light there at all. We may verify this by going behind the mirror to see. We shall later meet with images, known by contrast as *real* images, from which light actually does diverge.

We have taken the lamp S here to be a single point. It is easily seen, however, that if we have an extended object—as a lamp, of course, actually is—the same rule of reflection, applied to the light from each point of the object, gives us an image which is a replica of the whole object. In Fig. 10, for example, we can verify that G′H′ is the image of the arrow GH.

Mirrors

Bodies may be divided into two classes—those which scatter the light they do not absorb or transmit, and those which reflect it regularly according to the law just given. The

latter we call *mirrors* ; they are exemplified by smooth metal or liquid surfaces, the latter, however (mercury excepted), reflecting only a small proportion of the incident light. No mirror is absolutely perfect, or we should not see it at all, but only the image reflected in it. The slight imperfections in the surfaces of ordinary mirrors scatter enough light for us to see them when they are illuminated, from whatever direction we look towards them.

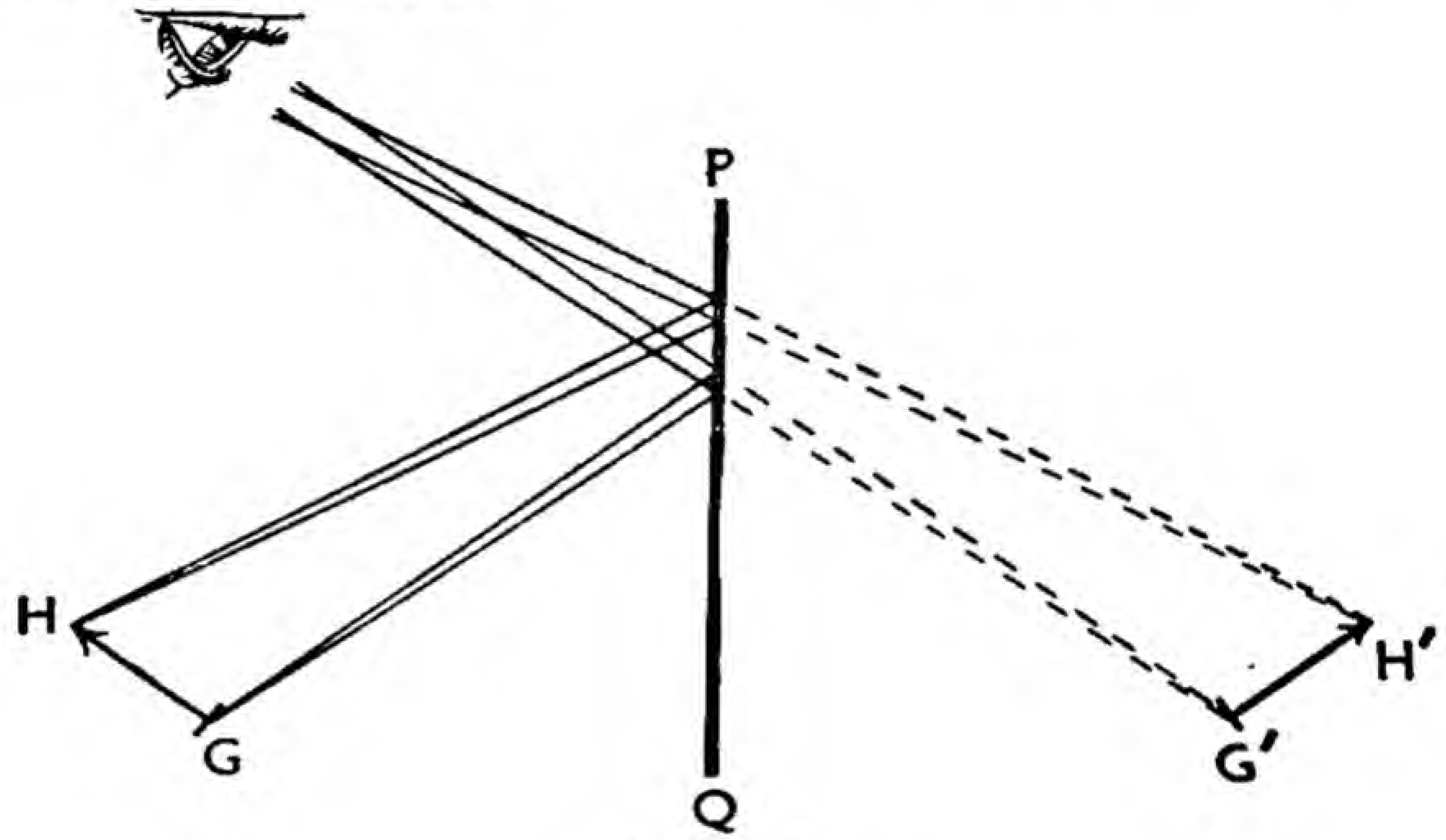


FIG. 10

Formation of Image of Extended Object GH in Mirror PQ

Since all bodies scatter some light, all that we have said about a lamp, which is itself a source of light, applies to any ordinary object, if it is illuminated by some external source. The arrow GH, for instance, in Fig. 10 need not be self-luminous. If it is lit up from without, the light falling on each point of it is scattered in all directions, so that it behaves just as though it were itself the source of that light. That is why, for example, we can see an image of ourselves when we look in a mirror in a lighted room.

Rotation of Mirrors: An interesting and important question can now be answered, namely—if I observe the image of an object in a mirror, and the mirror is then rotated through

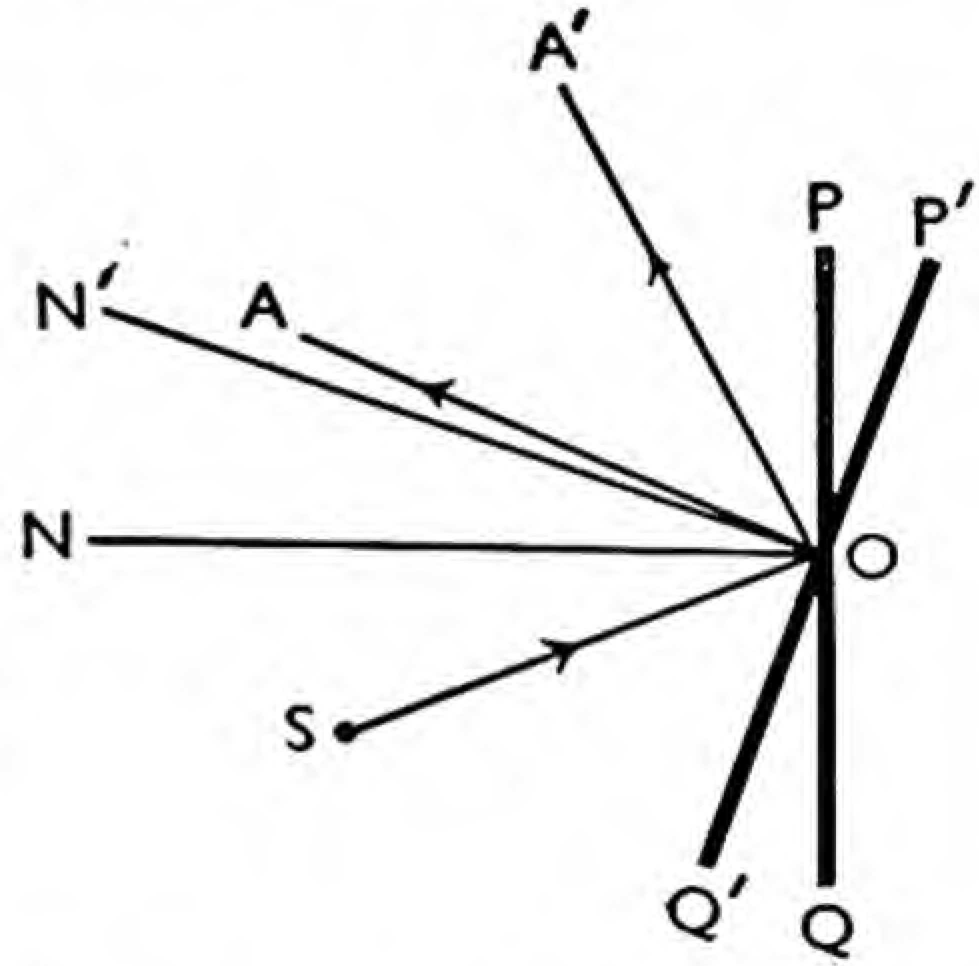
an angle, in what direction must I look to see the image by means of rays incident in the same direction as before? In other words, if the mirror is rotated, what happens to the reflected light? In Fig. 11 let the source of light be at S, and let PQ and A be the original positions of the mirror and the eye respectively. (Single rays are drawn instead of pencils, to avoid confusing the diagram). Then the angles SON and NOA are equal. Now let the mirror be rotated through a small angle α to the position P′Q′. The new

FIG. 11

Rotation of Reflected Ray when Mirror is Rotated

PQ, PQ′	Successive Positions of Mirror
ON, ON′	Successive Normals to Mirror
SO	Incident Ray
OA, OA′	Successive Reflected Rays

Rotation of Reflected Ray, AOA′ = twice Rotation of Mirror, POP′



normal will then be ON′, which will of course make an angle α with the old normal ON. The ray proceeding along SO will now have a new angle of incidence, $SON' = SON + \alpha$. The new angle of reflection, therefore, will be N′OA′, which, of course, will also be equal to $SON + \alpha$.

$$\text{Hence}\quad SOA' = 2SON' = 2SON + 2\alpha \quad . \quad . \quad . \quad (2{\cdot}1)$$
$$\text{But}\quad SOA = 2SON;$$
$$\text{hence}\quad AOA' = SOA' - SOA = 2\alpha \quad . \quad . \quad . \quad (2.2)$$

The reflected ray is therefore turned through twice as large an angle as the mirror. It follows, of course, that if the mirror is steadily rotated at a certain rate the reflected ray will rotate at twice that rate. It follows also that if the eye is kept fixed at A the direction of the new object which

is seen will make an angle with ON′ equal to AON′, and this direction will be inclined at 2α to SO. Thus we may say that the incident ray, for a constant direction of reflection, also moves through twice as large an angle as the mirror.

The Sextant

This fact is made use of in a number of optical instruments, of which the most important for our purpose is the *sextant.* This is an instrument for measuring the angle between the directions of two distant objects. Its construction is illus-

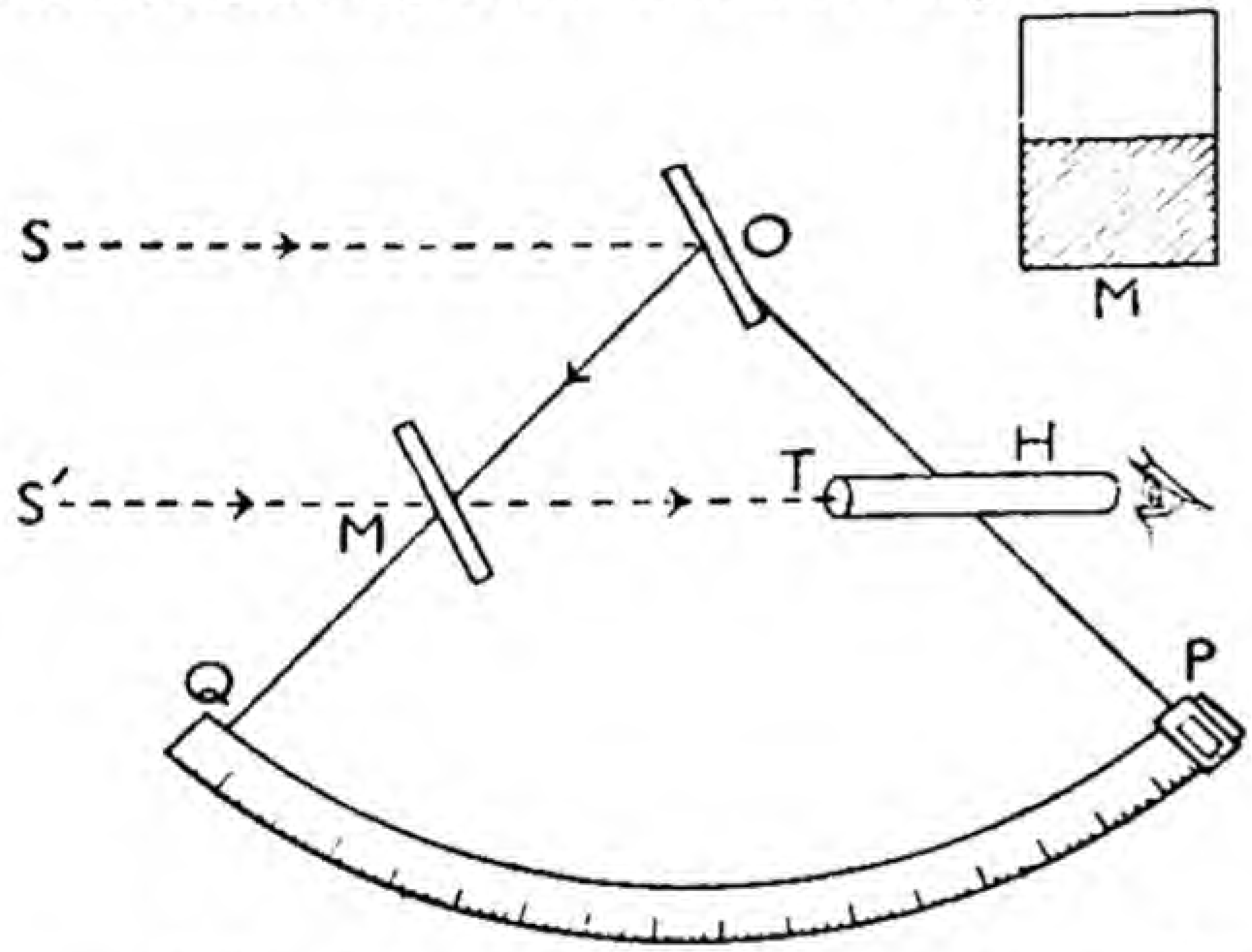


FIG. 12

The Sextant

In the position shown, two images of the same distant object, formed respectively by rays SOMT and S′MT, are seen superposed in the telescope H. As the arm OP, carrying the mirror O, rotates about O the image seen by the former path changes

trated by Fig. 12. PQ is an arc of a circle, finely graduated in small fractions of a degree. A rod OP, pivoted at the centre O and carrying a vernier at P, has rigidly attached to it a small plane mirror O, which, of course, rotates about O as P moves over the scale. A second rod lying along the radius OQ has a " half mirror " M firmly fixed to it (M is shown separately in Fig. 12—the upper half consists of clear glass, while the lower half is silvered), and this rod and mirror cannot be moved with respect to the scale. A tube, or small telescope, H, lies just above OP but detached from it ; this also is fixed with respect to the scale. Dark glasses can, if necessary, be placed before the mirrors, to protect the eyes when the Sun is being observed.

Now the mirror at M is fixed so that, when the reading

of the vernier at P is the zero of the scale, it is parallel to the mirror at O ; and the inclination of the mirrors then is such that a distant object in the direction of the telescope can be seen through it in two ways—first, by direct light coming through the clear upper half of M, and secondly, by light falling in the same direction on O and reflected therefrom to the lower half of M and thence into the telescope. The paths of the beams are shown by the dotted lines in Fig. 12, the light travelling, of course, from one mirror to the other along OM. The two views thus formed coincide with one another when the mirrors are parallel, and can be slightly separated by moving the arm OP slightly. The scale reading, when they are exactly coincident, should be zero, and if it is not, a correction must be applied to all readings to allow for this error. Now suppose that P is moved a short distance along the scale—rotating the mirror O through an angle α, say. The image seen in the telescope by light reflected from O will then be that of another object in a direction inclined at 2α to that of the first.

To measure the angle between the directions of two given objects, then, we have simply to get their images superposed in the telescope, and double the reading of the vernier at P. The scale is often graduated so as to make this doubling unnecessary ; for instance, an angle of $1°$ is marked as $2°$. The scale reading is then the actual angle required.

Uses of the Sextant : One of the chief uses of the sextant is to determine the altitude of the Sun—*i.e.* the angle between the direction of the Sun and that of the horizon. When the true horizon cannot be seen owing to the intrusion of buildings or other obstacles, the angle between the directions of the Sun and its image in a horizontal mirror is observed ; this is twice the Sun's altitude, as may easily be seen from Fig. 13, where S and S′ are the Sun and its image respectively, and O is the observer, effectively at the same level as the horizontal mirror PQ. The angle SOS′ is clearly 2SOP.

A great advantage of the sextant is that it does not need to be levelled, for both images move together, no matter how much the instrument moves about during the observations. For this reason it can be used with accuracy on board ship or in an aeroplane, where other instruments for measuring angles would be useless. The altitude of the Sun is important

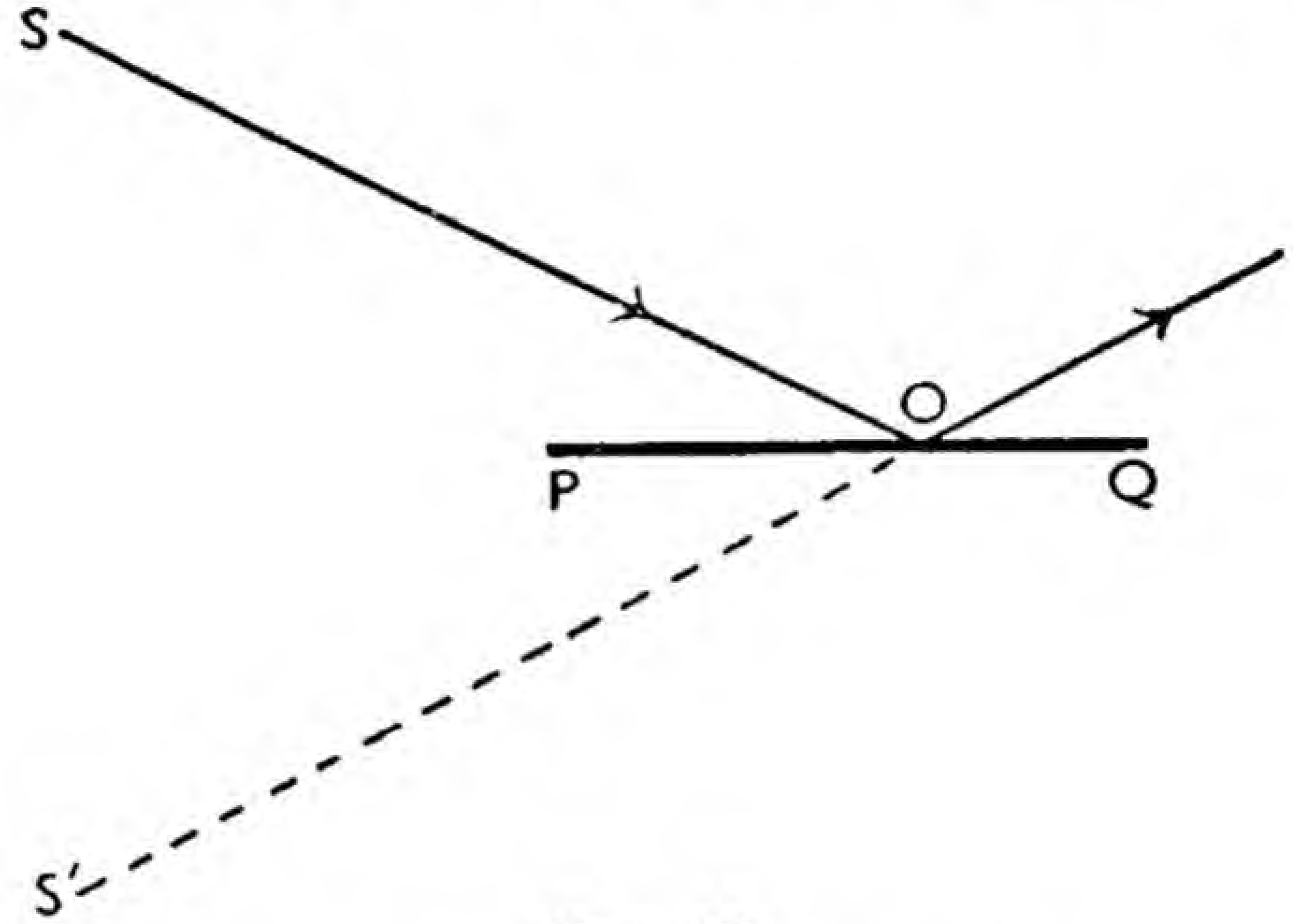


FIG. 13
Measurement of Altitude of Sun
PQ Horizontal Mirror S Sun S′ Image of Sun

in the calculation of latitude or local time, and it is constantly measured during sea voyages.

SPHERICAL MIRRORS

We have seen that the law of reflection of light ensures that all rays from a point object, after reflection from a plane mirror, proceed as though they came from another point—the image. If the surface of the mirror is curved, however, it does not follow that this will still be true. It can be shown, nevertheless, that if the surface is part of a sphere, and subtends only a small angle at the centre of the sphere, the reflected rays will still proceed as if they came from a point (or, strictly speaking, a very small area), so that we still have

images produced. We must now see where such images are located in relation to the object and the mirror.

Concave and Convex Mirrors

We have two cases to consider—namely, those in which the object is on the same side of the mirror as the centre of curvature, and on the opposite side. The reflecting surface, of course, must face the object in each case. These two cases are illustrated in Fig. 14. In the former, represented by (*a*), we have a *concave* mirror, and in the latter, represented by (*b*), we have a *convex* mirror. The angle subtended by the surface AB at the centre of curvature C is called the

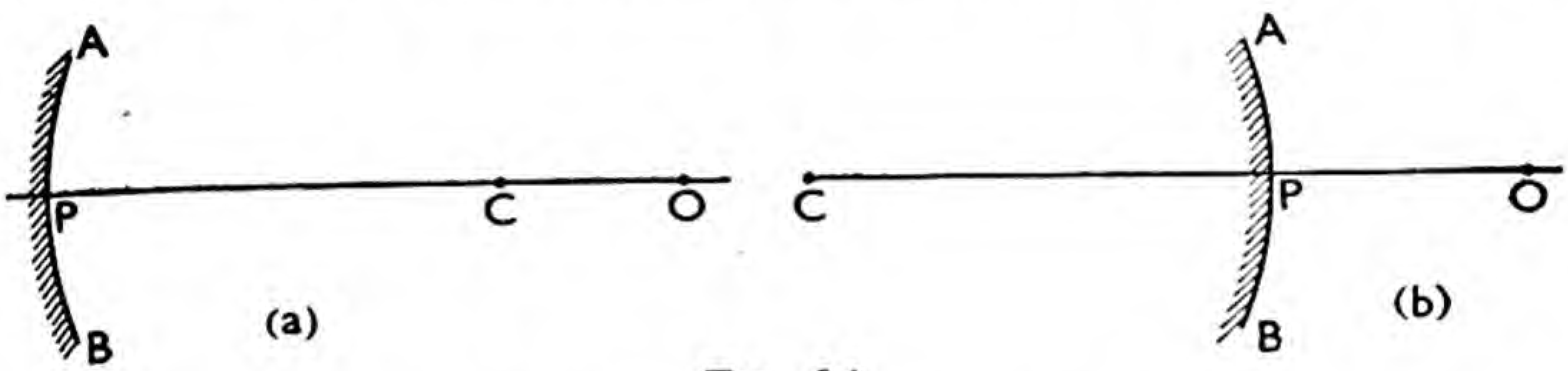


FIG. 14

Spherical Mirrors : (*a*) Concave, (*b*) Convex
P Pole C Centre of Curvature PC Axis O Object on Axis
Angle ACB = Aperture
The shading indicates the back of the mirror

aperture of the mirror, and the condition already mentioned for the formation of a point image may be expressed by saying that the aperture must be small. The central point of the surface P is known as the *pole* of the mirror, and the line joining P to C, the centre of curvature, is called the *axis*.

Image formed by Concave Mirror

Consider first a concave mirror, and suppose the object O is at a point on the axis. Then we can show that the image will also be at a point on the axis, the position of which depends on the position of O. It is obvious that if O is at C the image will be formed at C also, for, the surface being spherical, every ray from the centre meets it normally and is reflected

back along its own path. The reflected rays, therefore, all meet again at C and proceed outwards from that point. We have here an example of a *real* image, for the reflected rays actually meet there, and do not merely proceed as though they did.

Now suppose the object is farther from the mirror than C (Fig. 15). Then the ray travelling along the axis will meet the mirror normally and return along the axis, but another ray, meeting the mirror in a point M, say, will be reflected, by the ordinary law, so that the angles OMC and CMI are

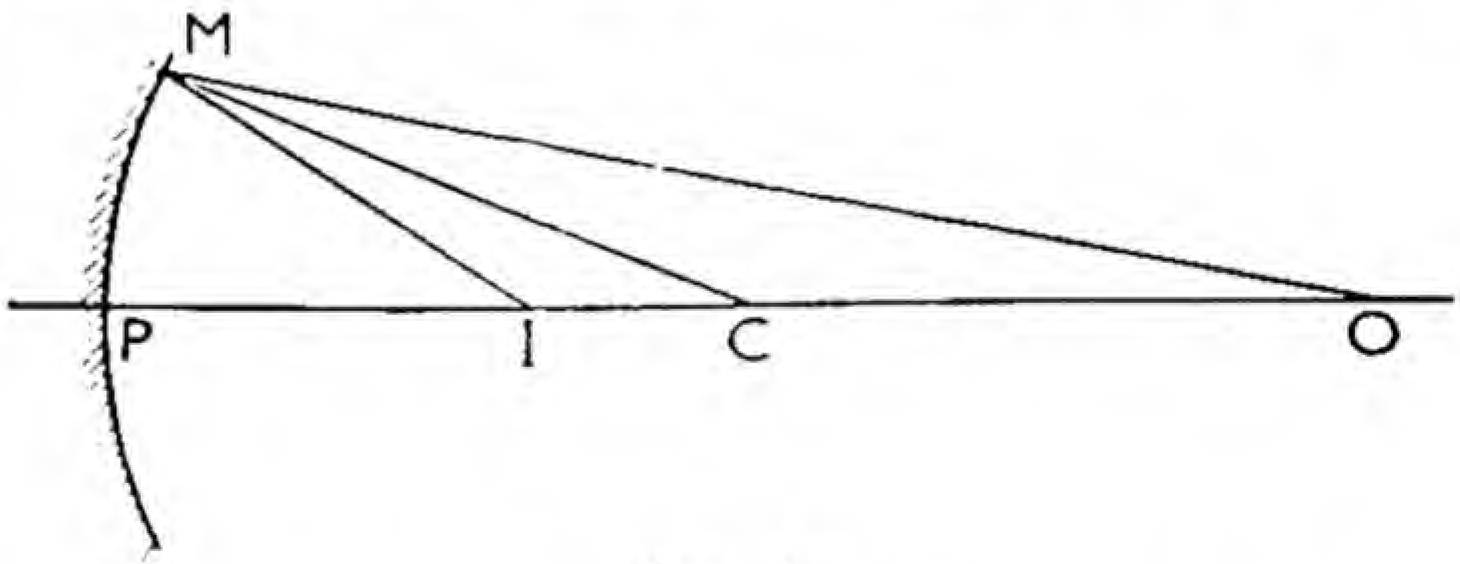


FIG. 15

Relative positions of an axial object and its image in a Concave Mirror
O Object I Image C Centre of Curvature of Mirror

equal, and will intersect the axially reflected ray at I. If, then, the other reflected rays pass through I, this will be the position of the image. We will show that, under the conditions assumed, they do so.

Since OMC = CMI, we have, by ordinary geometry,

$$\frac{CO}{CI} = \frac{MO}{MI} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2{\cdot}3)$$

Now, since the aperture of the mirror is small, M is close to P, so that $\frac{MO}{MI} = \frac{PO}{PI}$ very approximately.

$$\text{Hence} \quad \frac{CO}{CI} = \frac{PO}{PI} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2{\cdot}4)$$

$$\text{or} \quad \frac{PI}{CI} = \frac{PO}{CO} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2{\cdot}5)$$

Now $\frac{PO}{CO}$ is constant for a given mirror and a given position of the object. Hence $\frac{PI}{CI}$ is the same, whatever ray we take from O. That is to say, I is the same point for all rays, since there is only one point which divides PC into two parts bearing a given ratio to one another. Hence all rays from O proceed, after reflection, to I. They continue beyond it, of course, and the eye, placed so as to receive some of them, sees an image at I. This is shown for a small pencil of rays

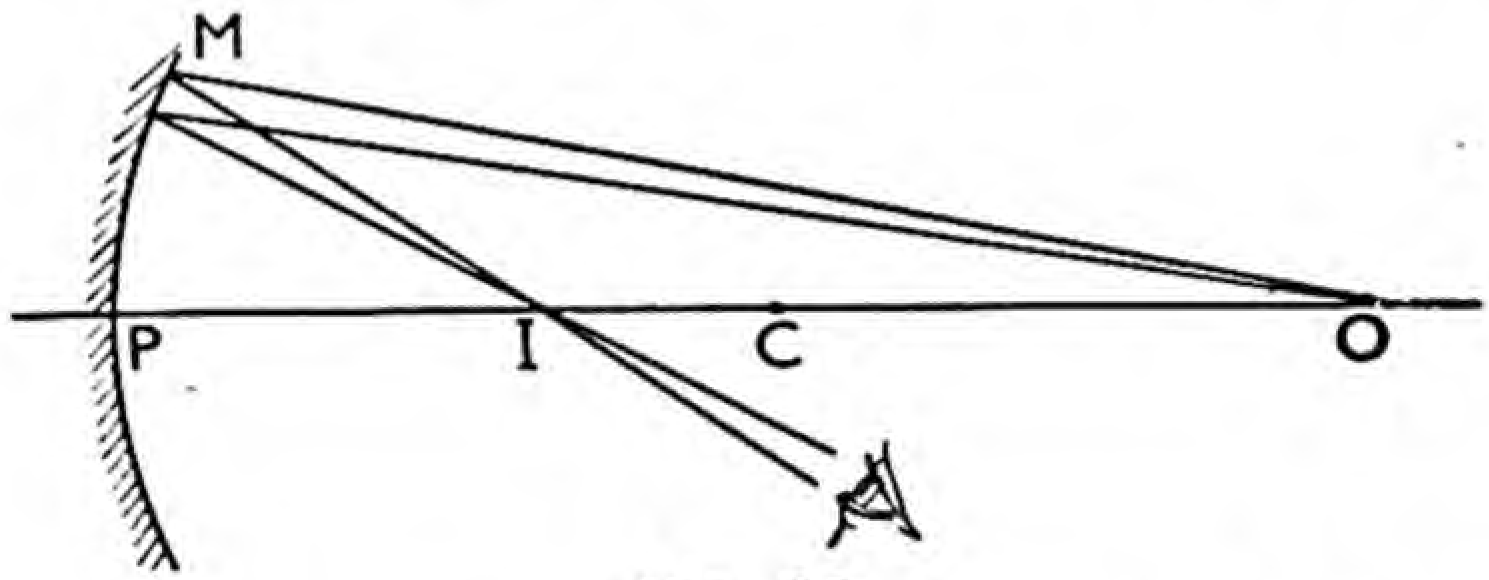


FIG. 16
Formation of Image I of Axial Object O by a pencil of rays incident on a Concave Mirror

in Fig. 16. A point such as I, towards which rays converge after reflection, is called a *focus* of the rays.

A real image may be received on a screen, and forms a picture of the object which can be seen from any position, since the screen scatters the light which falls on it. A screen, however, is not necessary to enable us to see the image, so long as we place our eye within the cone of light issuing from the mirror. Thus, in Fig. 17, if the eye is within the angle GIH, light from the image will enter it and the image will be seen. From such a point as Q, however, no image will be seen, but if a ground glass screen or a suitably inclined card is held at I, the light falling there will be scattered in all directions, and the image of the object may be seen from any point.

The Sign Convention

Conventional symbols are adopted to represent the positions of object and image; thus the distances of the object and image, respectively, from the pole of the mirror are represented by u and v, and the radius of curvature of the mirror is called r. Signs are given to these quantities according to the following rule: *All distances are measured from the surface of the mirror, those in the opposite direction to that of the incident light being reckoned positive, and those in the same direction negative.**

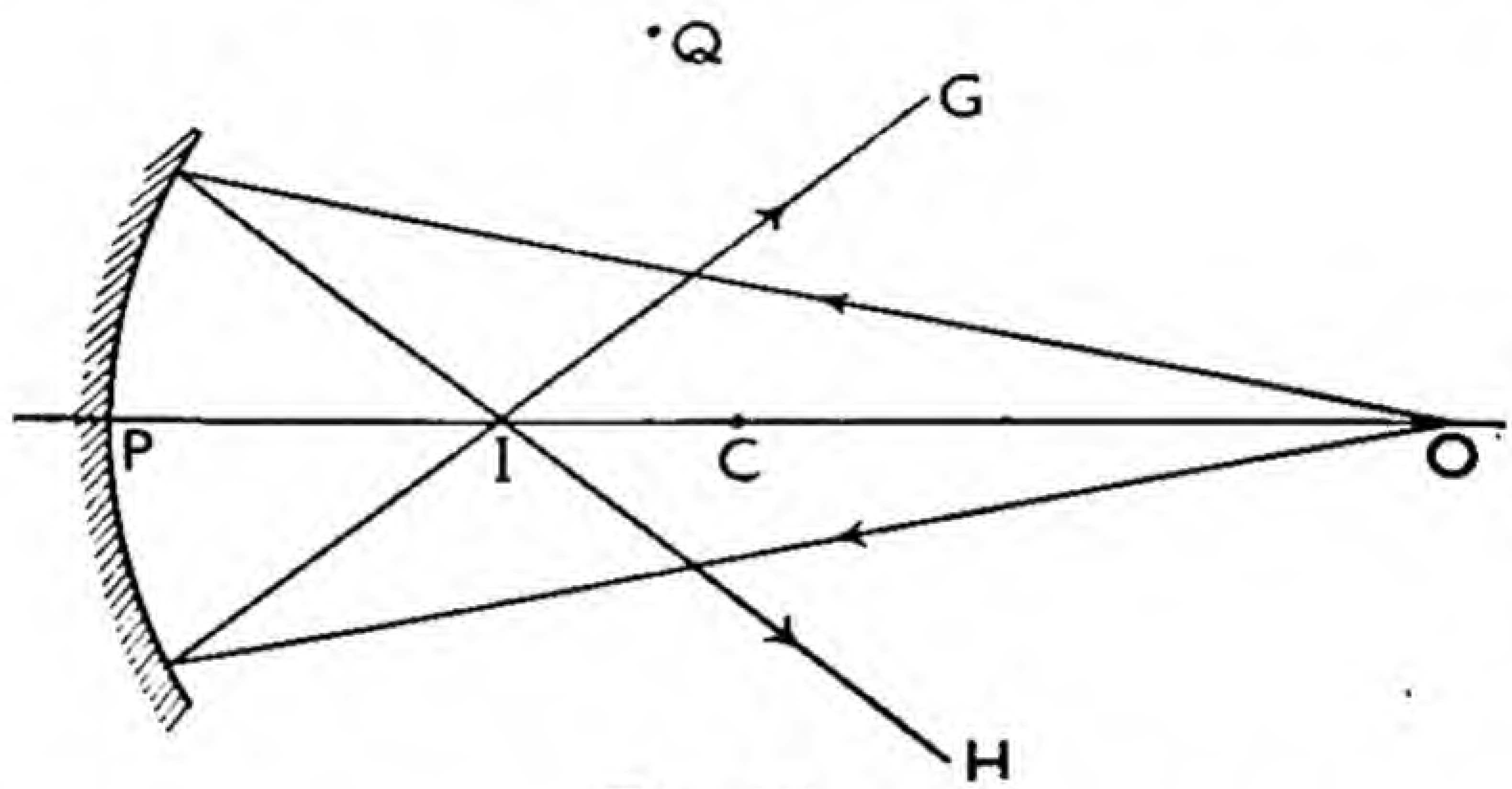


FIG. 17

Cone of Light GIH diverging from Image I of Axial Object O in Concave Mirror

In the case just considered we have, clearly, PI $= +v$; PC $= +r$; and PO $= +u$. The above equation (2·5) then becomes

$$\frac{v}{r-v} = \frac{u}{u-r}$$

which reduces to $\dfrac{1}{v} + \dfrac{1}{u} = \dfrac{2}{r}$ . . . . . . (2·6)

This is a general equation connecting the positions of object and image with the radius of the mirror.

* Alternative conventions are sometimes used, giving different forms of the resulting equations. To save confusion we give only one form, and keep to it throughout.

Principal Focus and Focal Length

A particularly important case occurs when the object O is very far from the mirror (" at infinity," as we say), so that the rays falling on the mirror are all approximately parallel to the axis.* The quantity $\frac{1}{u}$ is then negligibly small, and we have

$$\frac{1}{v} = \frac{2}{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2{\cdot}7)$$

so that $v = \frac{r}{2}$. The image is therefore half-way from the mirror to the centre of curvature. This point is called the *principal focus* of the mirror. Any image is, of course, at a focus, but the image formed by rays parallel to the axis is at the *principal* focus. The distance from the mirror to the principal focus is called the *focal length* of the mirror, and is generally represented by f. Our equation can therefore be written

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2{\cdot}8)$$

Positions of Object and Image

Let us see how the image moves as the object is brought from infinity towards the mirror. As we have seen, when u is infinite the image is at F—half-way between P and C (Fig. 18). As u decreases $\frac{1}{u}$ increases, so that $\frac{1}{v}$ must decrease to keep the sum equal to the constant $\frac{1}{f}$. Hence v increases —*i.e.* the image moves to the right. It is easily seen that as the object moves from infinity up to C the image moves from F to C—object and image coinciding, of course, at C. If the movement of the object is continued to the left of C,

* It is clear that the farther O is from the mirror, the more nearly parallel are the rays from it which fall on the mirror. Absolutely parallel rays may, therefore, be regarded as emanating from an object infinitely distant. Such an object would, of course, radiate in all directions like any other, but the particular pencil of its rays which fell on a finite surface would be parallel.

clearly the image moves farther to the right, for, from the law of reflection as well as from the form of equation (2·8),

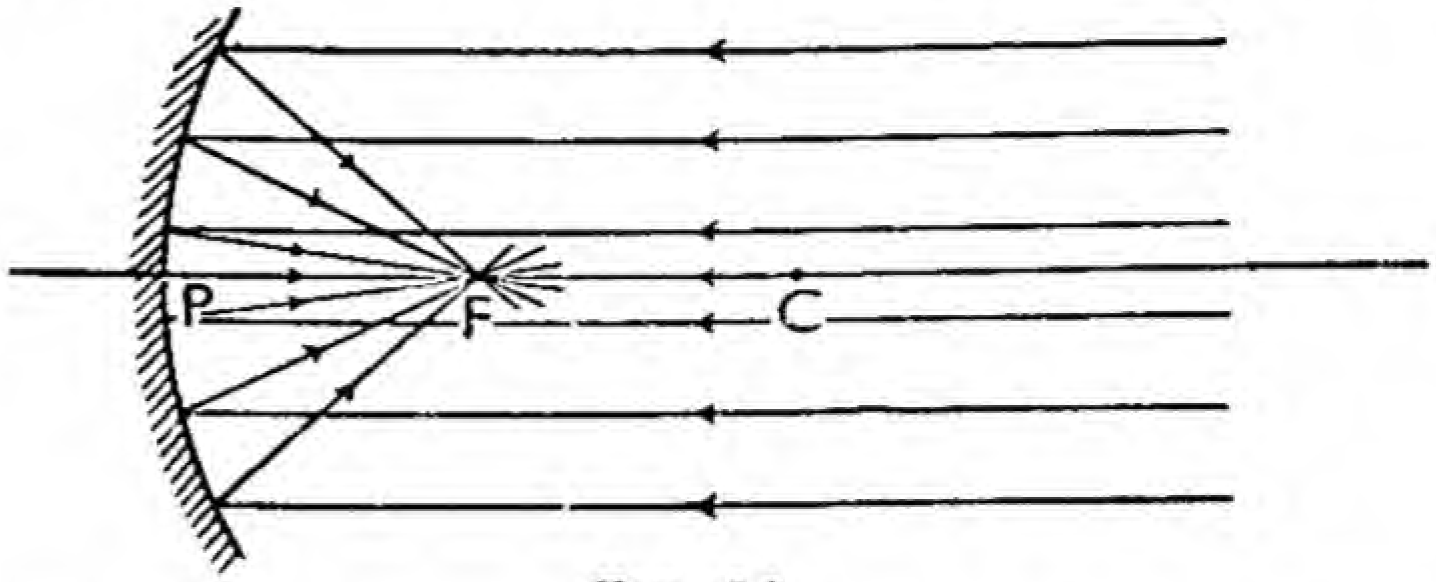


Fig. 18

Reflection of beam of light, parallel to the axis of a Concave Mirror, through the Principal Focus
P Pole C Centre of Curvature F Principal Focus

the object and image are interchangeable—*i.e.* if the object is at I the image will be at O. Hence, as the object moves from C to F the image moves from C towards infinity. (The points O and I, regarded as points on the axis, are

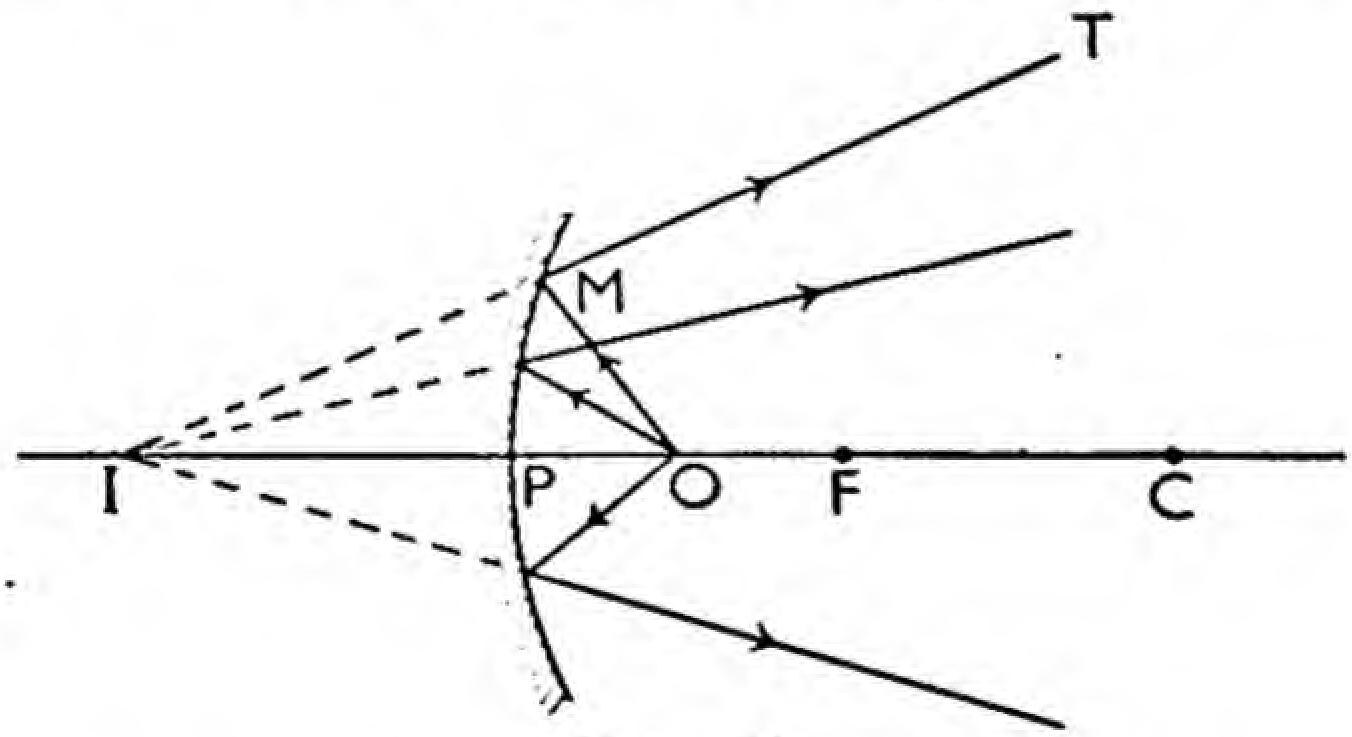


Fig. 19

Virtual Image I of Axial Object O inside Principal Focus F of Concave Mirror

called *conjugate foci*, since, if the object is at one, the image is at the other.)

What happens when the object is moved inside F? The image cannot go beyond infinity, but the construction (Fig. 19) shows that the reflected rays then diverge from the

axis, and proceed as though coming from a point I *behind* the mirror. The image therefore changes to a *virtual* one. Its position is given by our formula, which keeps us right with regard to signs, for we have

$$\frac{1}{v}=\frac{1}{f}-\frac{1}{u}=\frac{u-f}{uf}$$

$$i.e.\quad v=\frac{uf}{u-f} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2{\cdot}9)$$

Now if u is less than f, v is negative—*i.e.* it must be measured from the mirror in the *same* direction as the incident light,

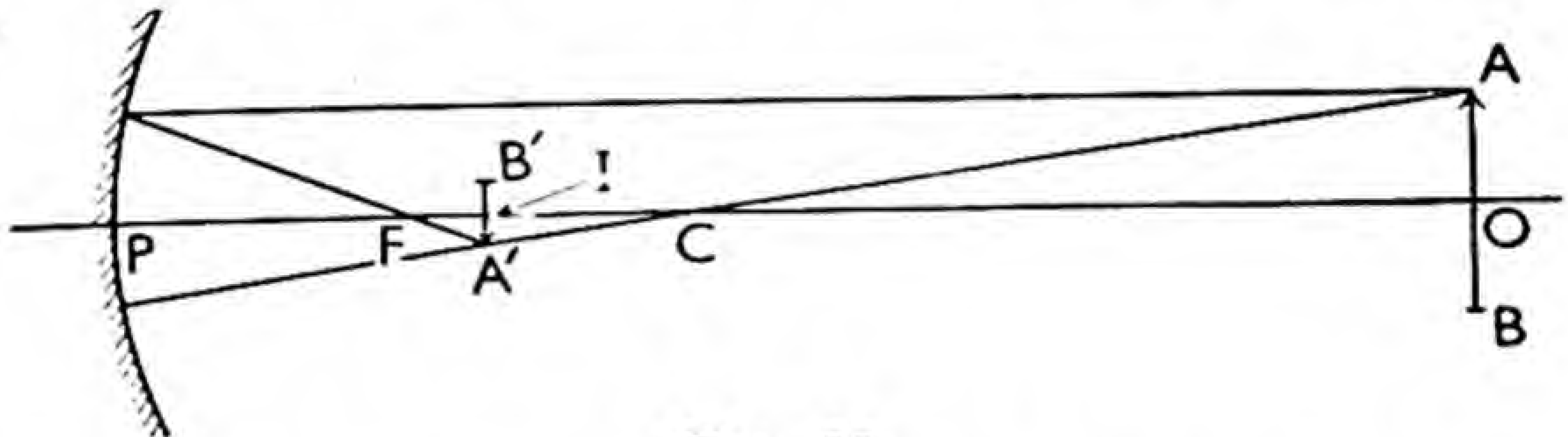


FIG. 20
Formation of Image A′B′ of extended Axial Object AB by Concave Mirror
C Centre of Curvature F Principal Focus

namely, towards the left. This gives us a point I behind the mirror. As u decreases $\frac{1}{u}$ increases, so that $\frac{1}{v}$ increases *numerically*, and v decreases. The image therefore moves from infinity on the left-hand side towards the mirror, and when the object actually arrives at the mirror the image arrives there also from the opposite side.

Objects of Finite Size

So far we have merely considered points on the axis, but we have essentially the same results with small objects of finite size, provided that they lie close to the axis. As an example, consider a small arrow situated as in Fig. 20. We may get an image of A by drawing two rays whose paths after

reflection we can easily determine. First, the ray through C will return along itself; and secondly, the ray parallel to the axis will proceed after reflection through F. These two rays intersect at A′, and, just as for axial points, all other rays from A pass after reflection through the same point, so that A′ is the image of A. The images of other points on the object are formed similarly, and we thus get an image A′B′ situated at a distance from the mirror given by equation (2·6) or (2·8) for axial objects.

It will be noticed that the image is inverted. This is so, as may easily be verified, so long as the object is farther from the mirror than F—*i.e.* so long as the image is real. Objects inside F form virtual images which, by drawing the construction, the student will easily find are erect.

We may get a simple expression for the size of the image relatively to that of the object. For, in Fig. 20, we have

$$\frac{\mathrm{OA}}{\mathrm{IA}'}=\frac{\mathrm{CO}}{\mathrm{CI}}=\frac{\mathrm{PO}}{\mathrm{PI}}\ (\text{as in }(2\cdot4))=\frac{u}{v} \qquad . \quad . \quad (2\cdot10)$$

Now $\frac{\mathrm{OA}}{\mathrm{IA}'}$ is the ratio of the sizes of corresponding portions of object and image, and is therefore the ratio of the sizes of the wholes. Hence we have the general rule that the sizes of object and image are in the ratio of their distances from the mirror. From this it can easily be seen that when the object is to the right of C it is larger than the image, and when between C and P it is smaller.

Convex Mirrors

In the same manner as before we can prove that a convex mirror will form images of axial points and small objects when the same conditions are fulfilled. It is clear, however, from Fig. 21 that, wherever the object may be, the rays from it which fall on the mirror diverge after reflection; so that the image is always a virtual one. The focus of rays parallel to the axis—the principal focus—is again half-way between

C and P, so that the focal length, f, is again $\frac{r}{2}$, but this time, of course, both r and f are negative, since PC and PF, measured from the mirror, are in the same direction as the incident light.

We can derive the same formulæ (2·6) and (2·8), but some care is necessary in observing the sign convention. We arrive without difficulty at the relation $\frac{\text{PI}}{\text{CI}} = \frac{\text{PO}}{\text{CO}}$ when all the lengths are measured without regard to sign—*i.e.* all are taken as positive.

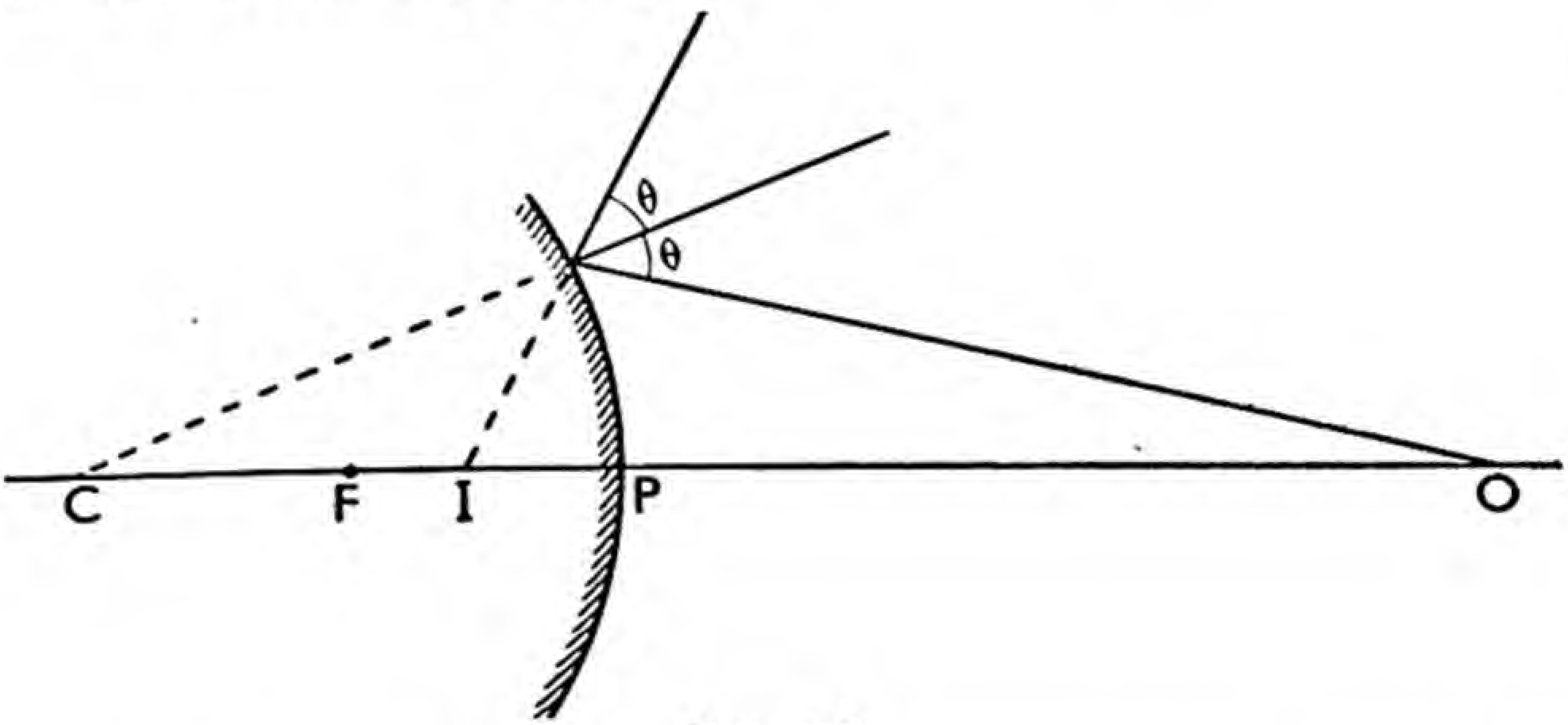


FIG. 21

Relative positions of an axial object and its image in a Convex Mirror
O Object I Image C Centre of Curvature F Principal Focus

We next have $\text{PI} = -v$; $\text{CI} = \text{PC} - \text{PI} = -r - (-v)$; $\text{PO} = u$; $\text{CO} = \text{PO} + \text{PC} = u + (-r)$.

$$\text{Hence} \quad \frac{\text{PI}}{\text{CI}} = \frac{\text{PO}}{\text{CO}}$$

$$\text{becomes} \quad \frac{-v}{-r+v} = \frac{u}{u-r}$$

$$\textit{i.e.} \quad \frac{1}{v} + \frac{1}{u} = \frac{2}{r} = \frac{1}{f} \quad . \quad . \quad . \quad . \quad (2{\cdot}11)$$

This is, therefore, a general formula for both concave and convex mirrors, but in applying it to any actual case we must remember to give the various quantities their proper signs.

The student may easily prove that, as the object moves from infinity leftwards towards the mirror, the image moves from F to the right towards the mirror. The image is, therefore, restricted to a very short region—namely, from F to P. It is always erect and diminished in size (and, as we have said, virtual), but reaches the same size as the object in the limiting case when the object actually reaches the mirror.

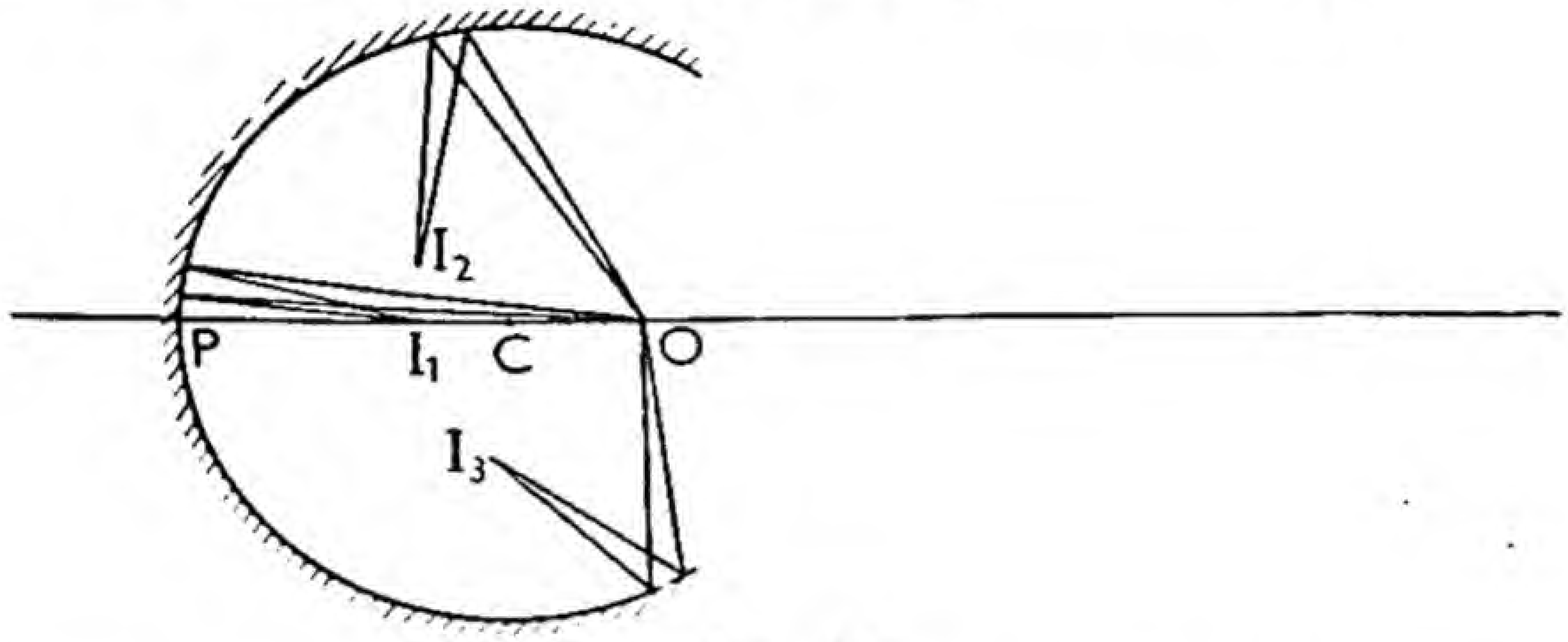


FIG. 22
Formation of Images I_1, I_2, I_3 by pencils of rays incident from an Axial Object O on different parts of a Concave Mirror

Mirrors of Large Aperture

If we have a mirror of large aperture and an axial point source, only the rays falling near the pole form an image on the axis. The rays forming a narrow pencil falling on the mirror some distance from the pole meet after reflection at some other point (which is, of course, also a focus), and the position of this point is different for different incident pencils. In Fig. 22, for example, three pencils are drawn, with the corresponding foci, I_1, I_2, and I_3. When the whole mirror is illuminated we accordingly have a whole series of images, lying along a curve called a *caustic*. The shape of this curve is shown in Fig. 23, in which the image for rays near the pole (the image I of Fig. 17) is at the cusp I. Such a curve can often be seen on the surface of tea in a teacup when sunlight falls on the inside of the cup. Here the source of light is not usually on the axis of the mirror, and the mirror

is cylindrical instead of spherical, so that the conditions are somewhat different, but the *locus* of images is of the same general type.

This characteristic of spherical mirrors (known as *spherical aberration*) is inconvenient when we wish to concentrate all the incident light at a single point—or, alternatively, when we wish to send the rays from a point source all in the same

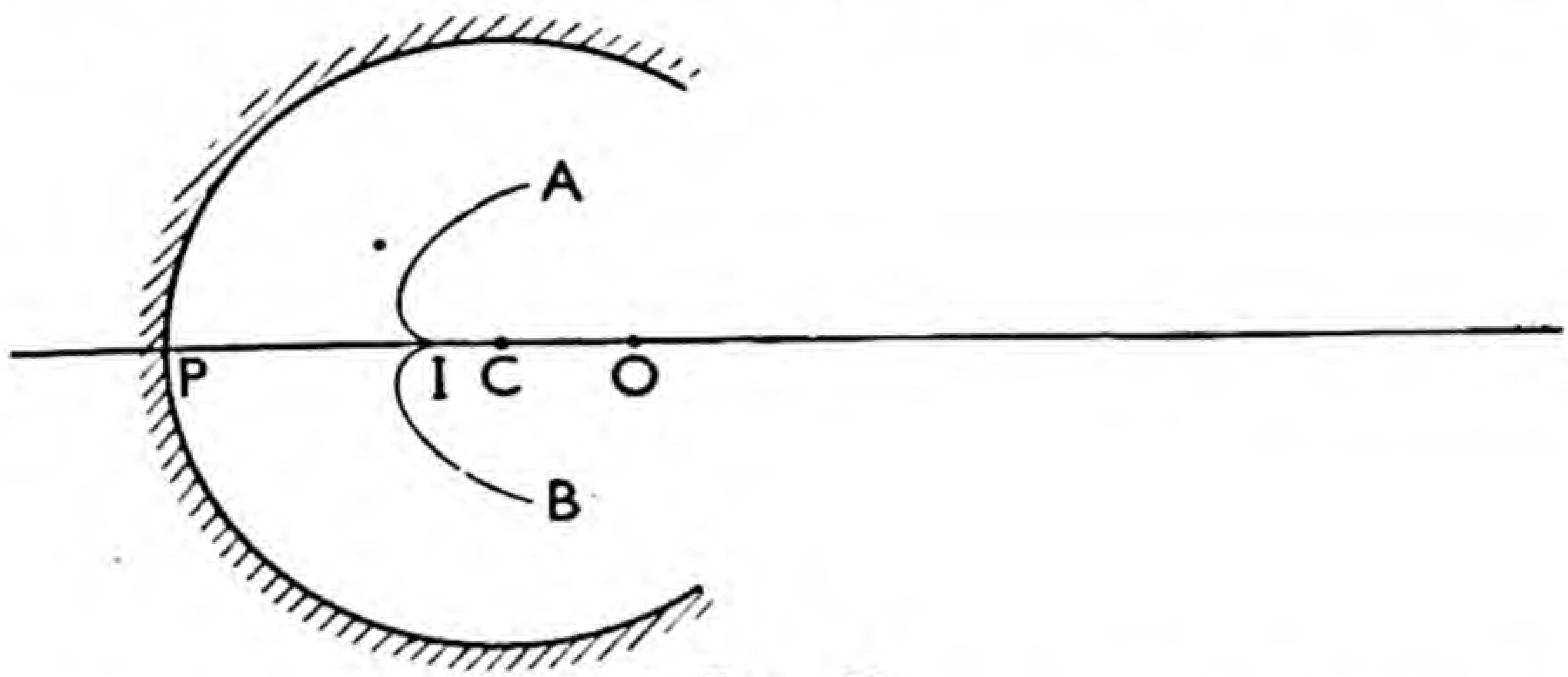


FIG. 23

Caustic AIB formed by reflection of rays from Axial Object O by Concave Spherical Mirror of wide aperture

I Position of Image formed by rays incident near the Pole P

direction. For example, in a searchlight mirror we place the source at the principal focus in order to concentrate the reflected light into a long parallel beam. Spherical aberration, however, prevents this with spherical mirrors, for the light from the edges of the mirror is not reflected parallel to the axis, as is that from the neighbourhood of the pole. A paraboloidal mirror, however, has the property of sending *all* rays from the focus along a beam parallel to the axis, and such mirrors are therefore used for this purpose.

EXERCISES

1. Describe experiments showing that light travels approximately in straight lines in a single transparent medium.

2. Explain the difference between scattering and regular reflection of light by a material surface. In what circumstances is an image of the source of light produced ?

3. State the law of reflection of light, and prove that the image formed by a plane mirror is as far behind the mirror as the object is in front.

4. Describe the sextant, and explain how you would use it to measure the altitude of the Sun.

5. Prove formula (2·11) for both convex and concave mirrors, explaining clearly the convention by which signs are given to the quantities occurring in it.

An object 1 cm. high is placed on the axis of a convex mirror 10 cm. from the pole. If the focal length of the mirror is 10 cm., find the position and size of the image, stating whether it is real or virtual, erect or inverted.

6. What is the spherical aberration of a mirror ? By actual measurement, using the law of reflection, draw the curve on which lie the images of a point on the axis of a large aperture concave mirror of 2 in. radius, the point being distant 3 in. from the pole of the mirror.

CHAPTER III

REFRACTION OF LIGHT

REFRACTION AT PLANE SURFACES

WHEN light passes from one transparent medium into another —*e.g.* from air into glass or water—it does not cease to travel in straight lines, but its direction is in general changed. Everyone knows that a stick partly immersed in water appears bent at the point of entry. The reason is that the light issuing from the immersed portion changes direction on coming out into the air, and since the object appears to us to be in the direction of the light which enters our eye, it appears that the stick has been bent. Thus, in Fig. 24, a pencil of light from A enters the eye as though it came from A′, and that is where the end A appears to be. The whole stick appears as though it took the course BCA′.

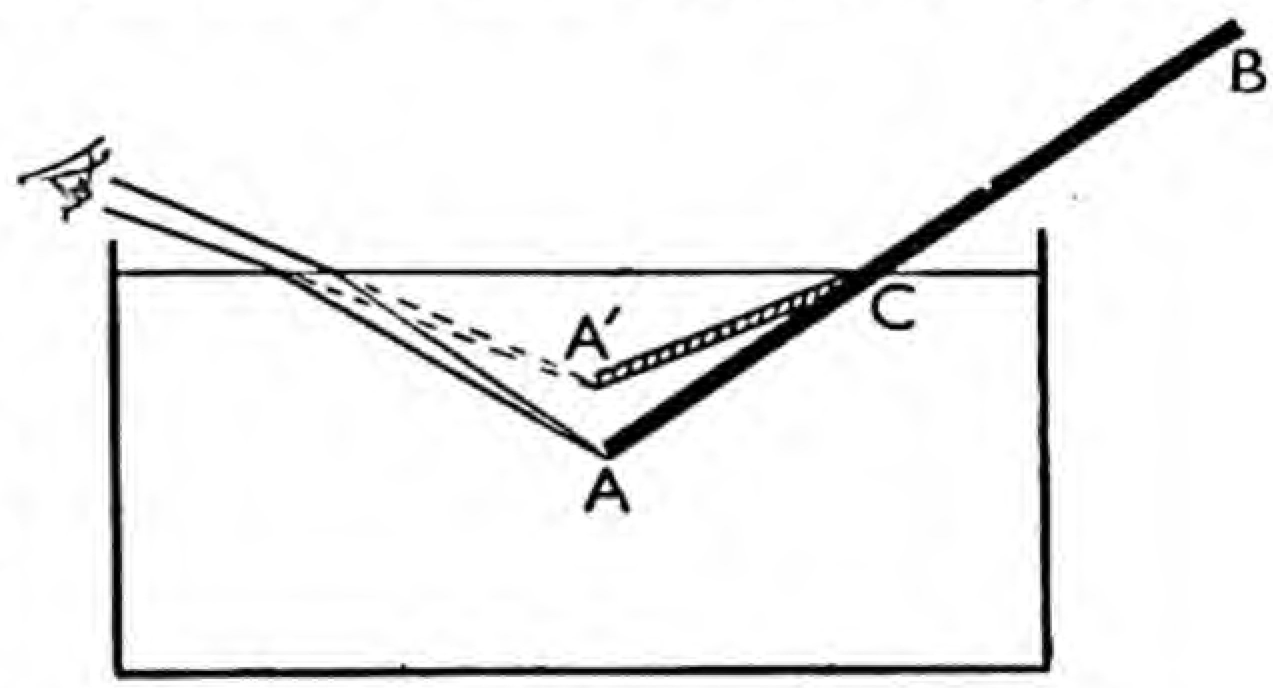


FIG. 24
Apparent bending of stick partly immersed in liquid
ACB Actual Course of Stick
A′CB Apparent Course of Stick

Refractive Index

This bending of light is called *refraction.* It takes place according to a rule just as definite as that of reflection. Calling the angles which the incident and refracted rays make with the *normal* to the surface (not with the surface itself) the *angle of*

incidence and the *angle of refraction*, respectively, the rule is that, for a given pair of transparent media, *the incident and refracted rays and the normal to the surface at the point of incidence lie in the same plane, and the sine of the angle of incidence, i, bears a constant ratio (known as the refractive index, μ) to the sine of the angle of refraction, r.* This is known as *Snell's law*. In symbols it is

$$\mu = \frac{\sin i}{\sin r} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3 \cdot 1)$$

The law is true for any pair of transparent media, but the refractive index is different for different pairs. It

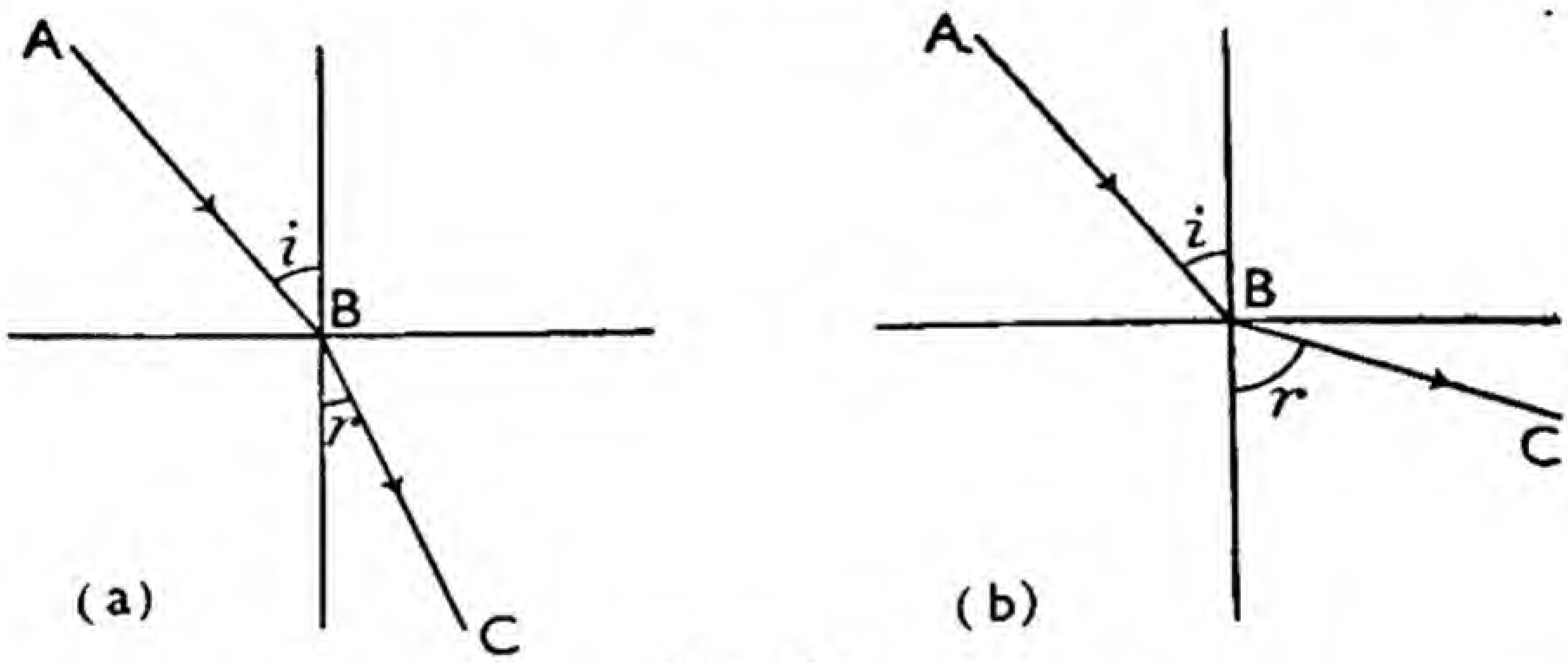


FIG. 25

Refraction of ray AB passing (*a*) from rarer to denser medium; (*b*) from denser to rarer medium

i Angle of Incidence *r* Angle of Refraction

varies also with the colour of the light for a single pair of media—*i.e.* with the wave-length according to the wave theory. For the present, however, we shall ignore this, and suppose that we are dealing throughout with light of one colour only.

Rare and Dense Media

If the refractive index is greater than 1, the second medium is said to be *optically denser* than the first. In that case it is clear that the ray is bent towards the normal (see Fig. 25(*a*)) on entering the second medium. If the refractive index is

less than 1, the second medium is said to be *optically rarer* than the first. The ray is then bent away from the normal on entering the second medium (Fig. 25 (*b*)). We find by experiment that if the light is reflected back along its own path in the second medium, it retraces its course when it returns to the first medium also. Thus, if (Fig. 25) a mirror is placed at C so that the light meets it normally, the light is reflected back along CBA. In stating the refractive index for two media we must therefore state the direction in which the light goes. This is often done by means of suffixes; thus ${}_1\mu_2$ indicates the refractive index when light passes from medium 1 to medium 2. From what we have said it is clear that

$$ {}_2\mu_1 = \frac{1}{{}_1\mu_2} \quad . \quad . \quad . \quad . \quad . \quad (3\cdot 2) $$

for, from the definition we have

$$ {}_1\mu_2 = \frac{\sin i}{\sin r}; \qquad {}_2\mu_1 = \frac{\sin r}{\sin i} \quad . \quad . \quad . \quad . \quad (3\cdot 3) $$

For simplicity, the term "refractive index" is usually restricted to the case where the medium from which the light is incident is a vacuum—the rarest medium known. The refractive index of every material medium is then greater than 1. Since air is a much more frequently occurring medium than a vacuum, and the refractive index of any other medium with respect to it is almost the same as for a vacuum,* we often substitute air for a vacuum and define the refractive index of a substance as the ratio of the sines of the angles of incidence and refraction when light passes into it from air. We then simply write the refractive index as μ, without the suffixes. As examples of the refractive indices of common

* Not quite, however. An interesting effect of the refraction of air is that we see the Sun before it rises and after it sets. When the Sun is a little below the horizon its light is bent by the atmosphere into a horizontal direction, and the Sun, therefore, appears to us to be in that direction.

transparent substances, it may be mentioned that the values for ordinary glass and water are respectively about 1·5 and 1·33. Glasses can be made, however, covering a fairly wide range of refractive index.

Image formed by Refraction

Let us now consider how an object appears when viewed through a parallel-sided block of glass, say. Each ray, incident at an angle i on the block, is bent towards the normal, making with it a smaller angle r in the block (Fig. 26), according to the fundamental relation (3·1), $\frac{\sin i}{\sin r}=\mu$. On emerging into the air it is bent again into its original direction, for the angle of incidence at B on the glass/air surface is equal to r, so that the angle of refraction into the air will be i. The ray, therefore, does not suffer a permanent change of direction by its passage through the glass, but it is moved laterally from the dotted course which it would have followed if the glass had not been there. The amount of lateral displacement increases with the angle of incidence. The ray incident normally on the glass ($i=0$) will go straight through without bending, for when $\sin i=0$ we must have $\mu \sin r=0$, *i.e.* $r=0$; hence this ray will suffer no lateral displacement.

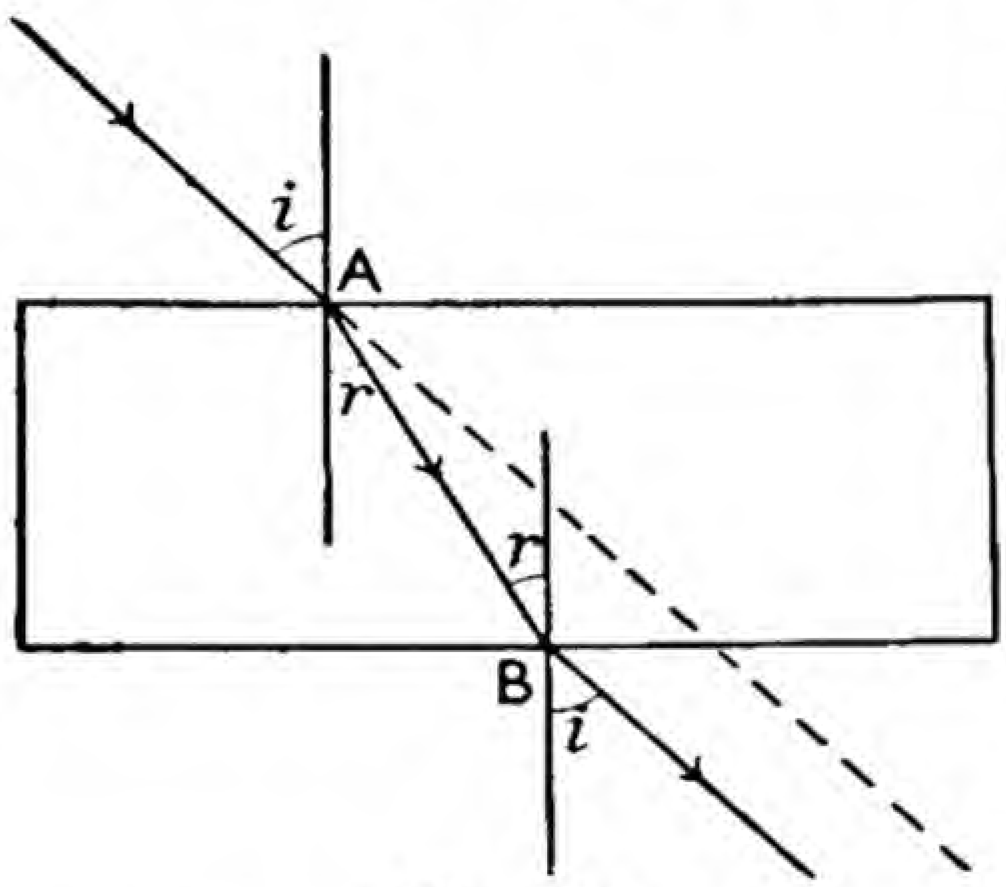


FIG. 26
Lateral deviation of ray refracted by parallel-sided Block of Glass

When we look at the object through the glass the eye takes in a small pencil of the light which the object is scattering in all directions; the particular pencil received, of course, depends on where the eye is placed. In Fig. 27 we have taken two positions. In each case an image I is formed, which is clearly a virtual one, and is nearer than the object

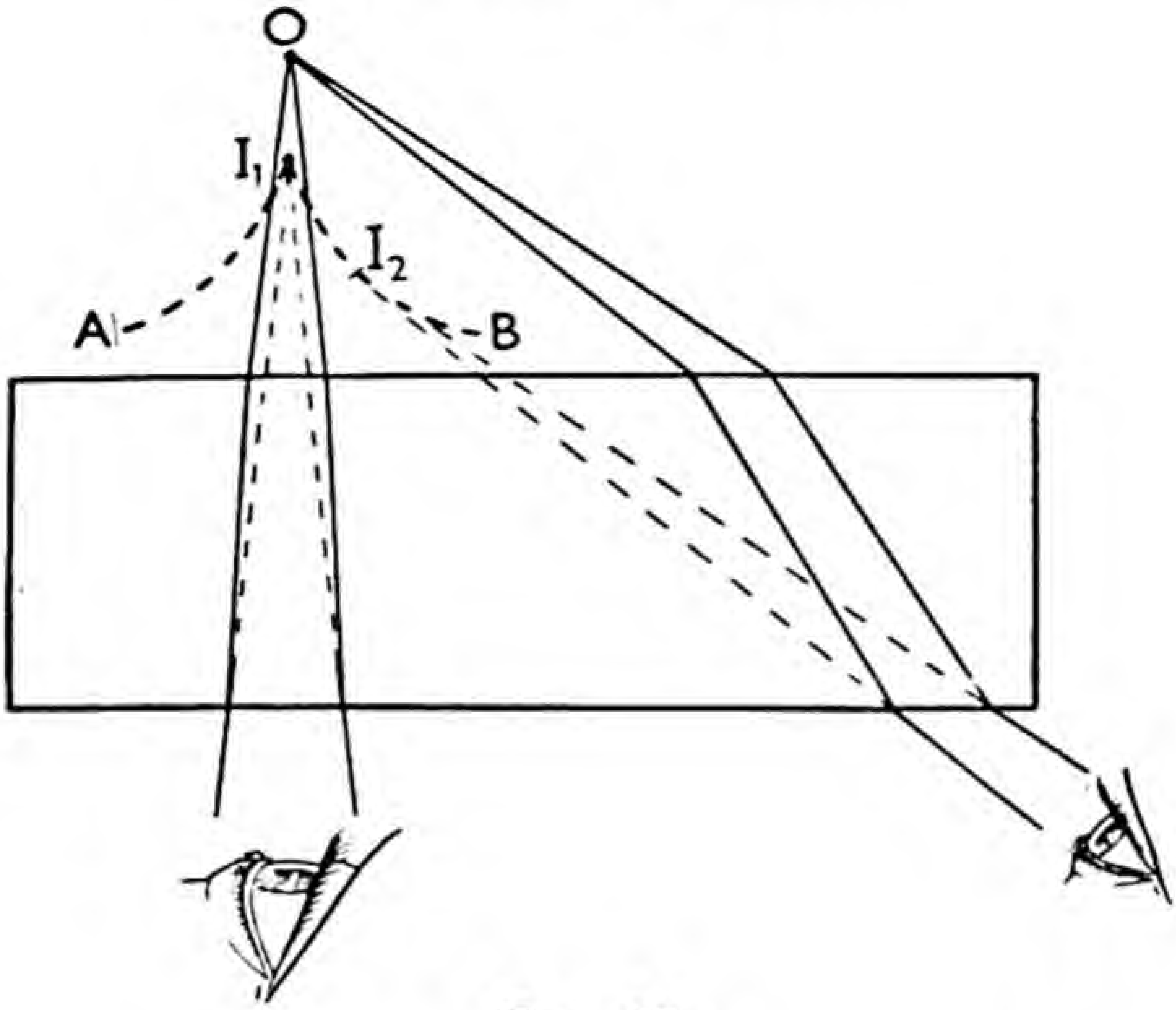


FIG. 27

Images of Object O formed by refraction through parallel-sided Block of Glass

I_1 Image formed by Normal Pencil

I_2 Image formed by Oblique Pencil

AI_1B Caustic on which lie Images produced by all Pencils

to the eye. The various positions of I, as the eye is moved, lie on a caustic curve similar to that obtained by reflection from a curved mirror of large aperture. It is shown as a dotted cusped curve in Fig. 27.

Refraction by Successive Media

It is important to know what happens to a ray of light when it passes through a succession of different media. Suppose we have two parallel-sided blocks made of different transparent materials, and a ray of light passes from air through them and into the air again (Fig. 28). Let air be called medium 1, and the other substances, media 2 and 3 respectively. Then we know from experiment that the light finally issues in the same direction as that in which it began, so that if

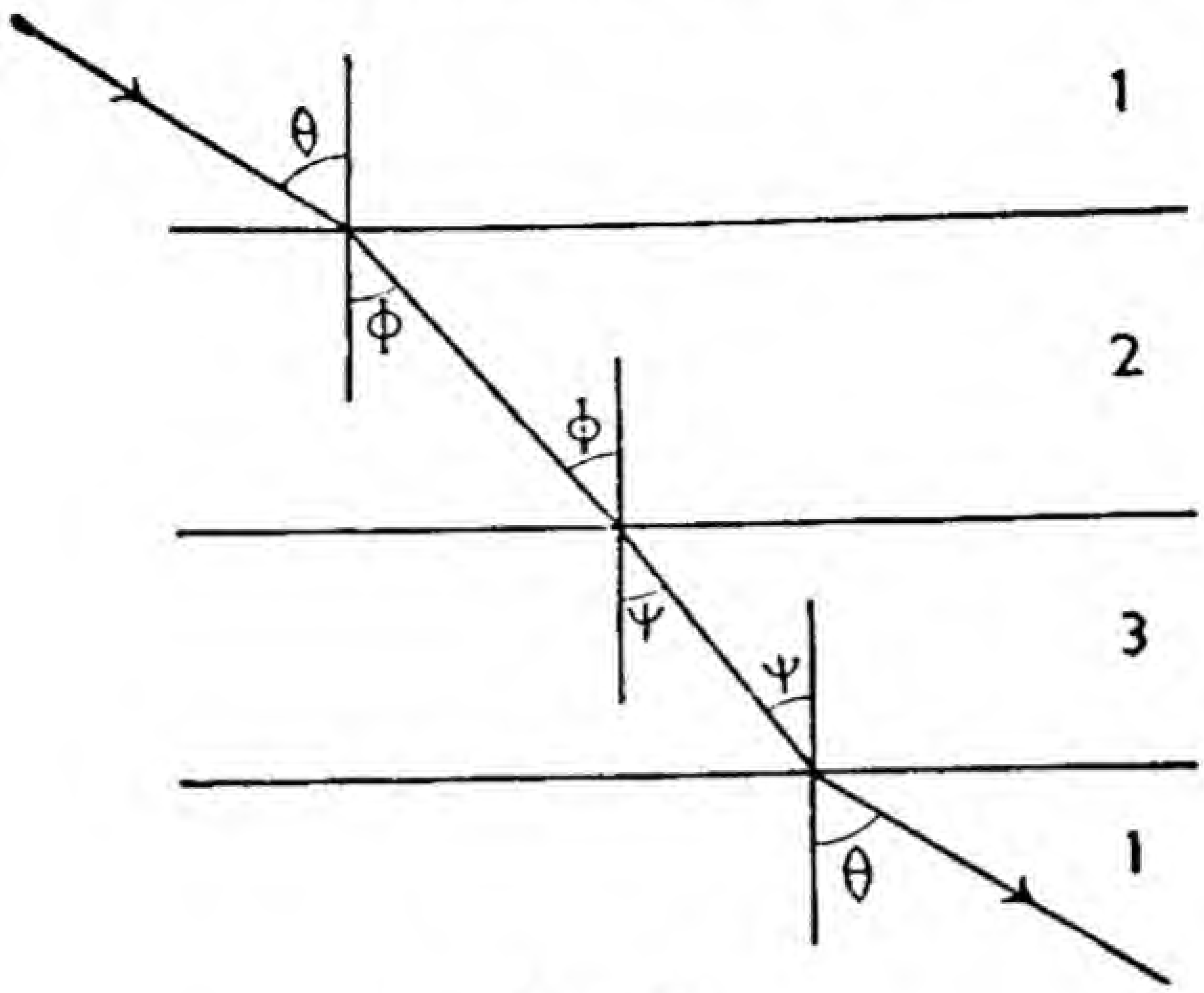


FIG. 28
Refraction of Ray by Successive Parallel-sided Media

the angle of incidence is θ, the final angle of emergence is θ also. We have then

$$_1\mu_2 = \frac{\sin\theta}{\sin\phi};\ _2\mu_3 = \frac{\sin\phi}{\sin\psi};\ _3\mu_1 = \frac{\sin\psi}{\sin\theta} \quad . \quad . \quad (3\cdot4)$$

$$\text{hence} \quad _1\mu_2 \times {_2\mu_3} = \frac{1}{_3\mu_1} = {_1\mu_3} \quad . \quad . \quad . \quad . \quad (3\cdot5)$$

In the same way, by taking more media we can show that

$$_1\mu_2 \times {_2\mu_3} \times \ . \ . \ . \ \times {_{n-1}\mu_n} = {_1\mu_n} \quad . \quad . \quad . \quad (3\cdot6)$$

Since the first medium is air, we can drop the suffix 1, and call μ_2 and μ_3 the refractive indices of the other media. Our result then becomes

$$_2\mu_3 = \frac{\mu_3}{\mu_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3\cdot7)$$

For example, the refractive index when light passes from water into glass is

$$\frac{\text{refractive index of glass}}{\text{refractive index of water}}$$

Multiple Images in Mirror

Most people have noticed that when a light is held in front of an ordinary mirror a series of images is seen. The formation of these is explained as follows. In Fig. 29 ABCD represents a section of the mirror, CD being the silvered surface. A pencil of light incident from O is partly reflected and partly transmitted at AB. The reflected pencil forms a virtual image at I_1, while the transmitted (and, of course, refracted) portion is wholly reflected at the silvered surface and is then partly refracted into the air at AB and partly reflected back into the glass. The refracted portion appears to come from a second image, I_2, while the part reflected back repeats this behaviour, and so a series of images is formed, gradually getting weaker as the light is used up.

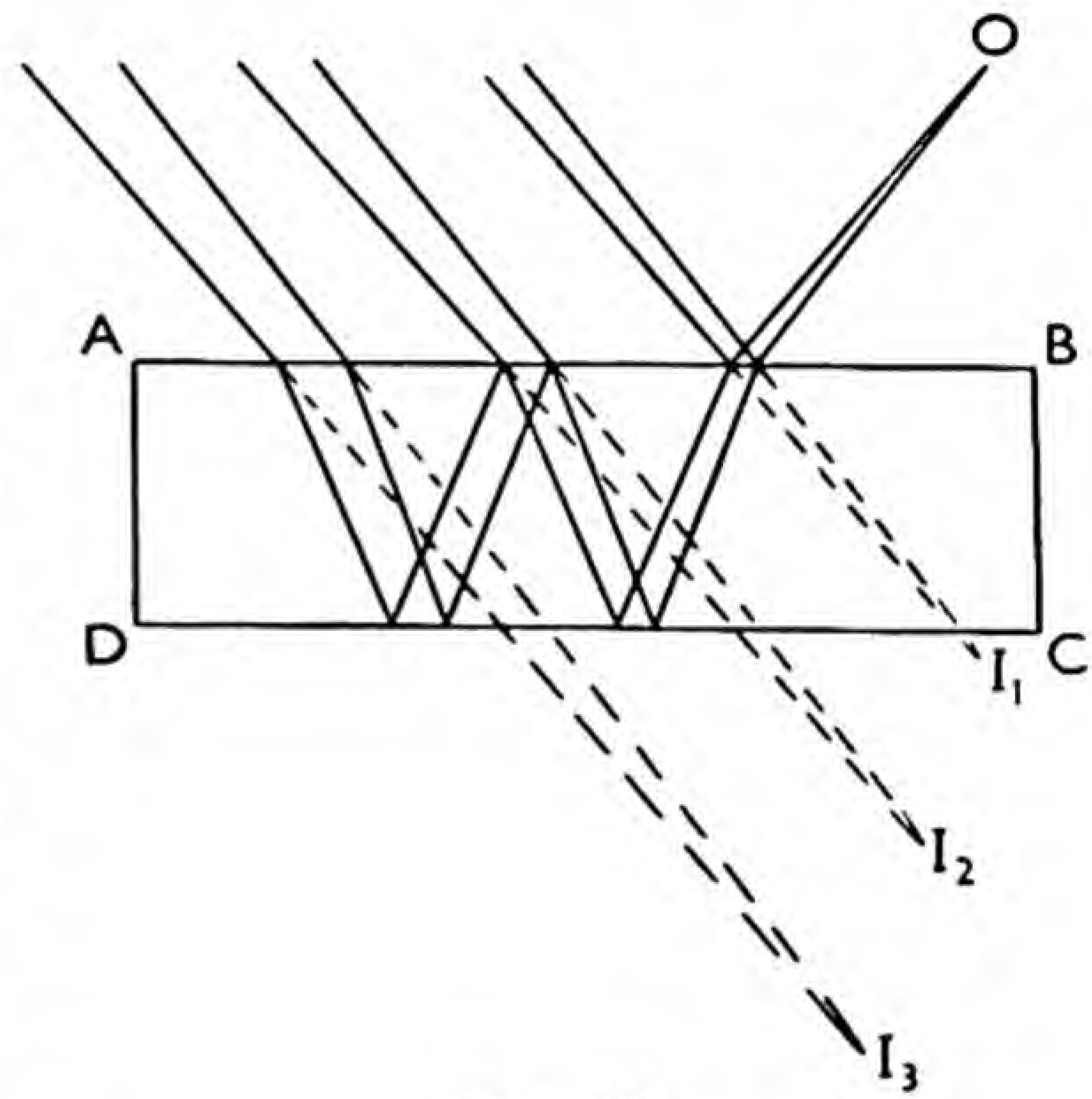


FIG. 29
Formation of Multiple Images by a Thick Mirror
O Object I_1, I_2, I_3 Images

Most of the light goes into I_2, since the silvered surface has a high reflecting power and the clear glass surface a high transmitting power.

It is well to warn the student against a false description of this process, which is very commonly given, in terms of single rays instead of pencils of light; it is illustrated in Fig. 30. If this represented what actually occurred, we would see not several images but only one, for all the rays entering the eye, and shown as issuing from different images, are

parallel, *i.e.* they all come from the same direction and therefore apparently from the same object. Indeed, if we were to make all the light falling on the glass parallel by placing O at the focus of a paraboloidal mirror and allowing only the reflected rays to fall on the glass, we should in fact see only one image, provided the surfaces of the glass were accurately parallel. To get the series of images we must have a diverging incident pencil, and the images are found at the points from which the secondary diverging pencils appear to come.

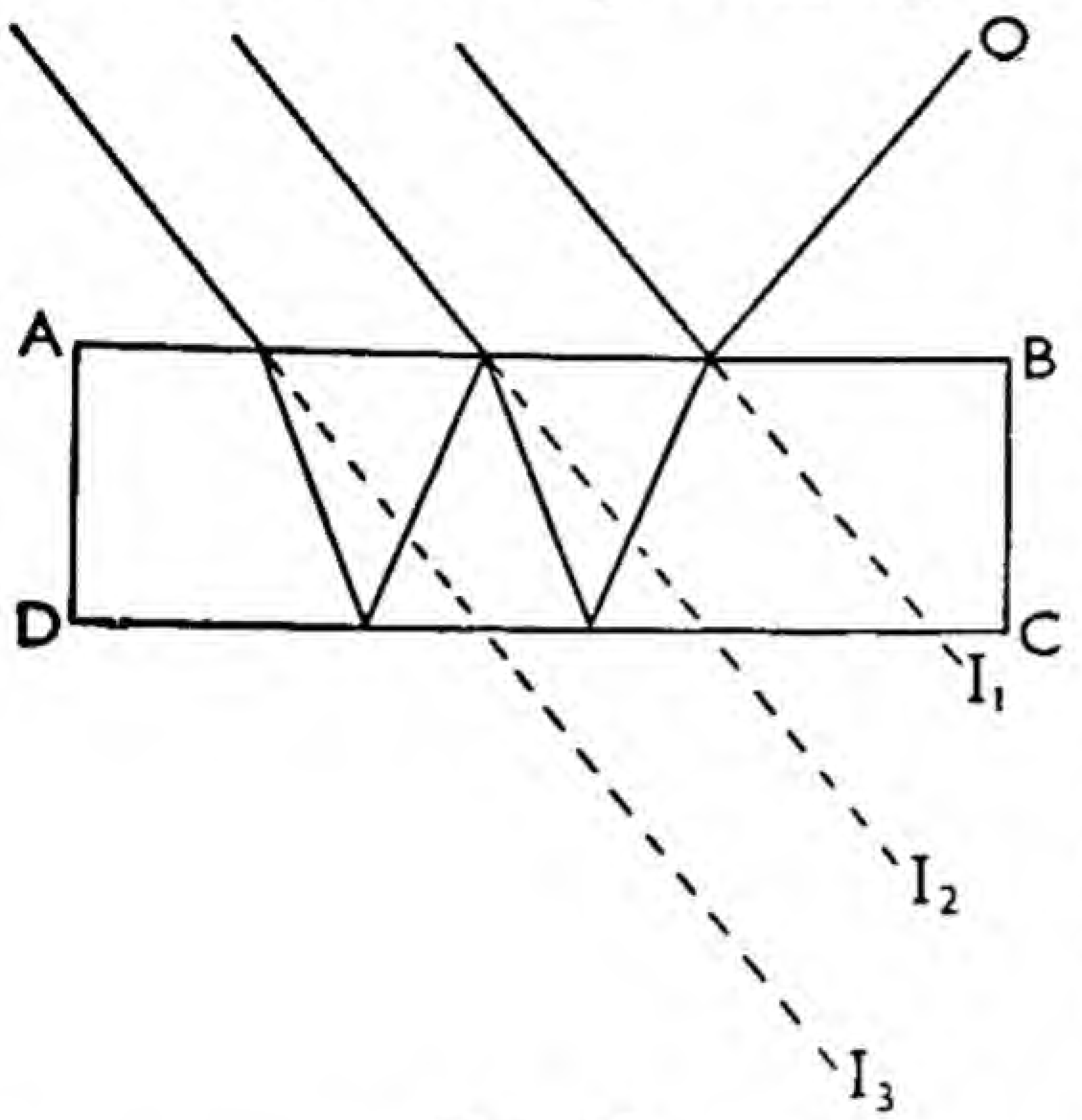


FIG. 30
False diagram of Formation of Multiple Images by a Thick Mirror

Total Internal Reflection

Suppose an object is placed beneath a glass block, or at the bottom of a basin of water. The light by which it is seen by the eye in air is refracted at the surface, and specimen rays are shown in Fig. 31. Glass or water being optically

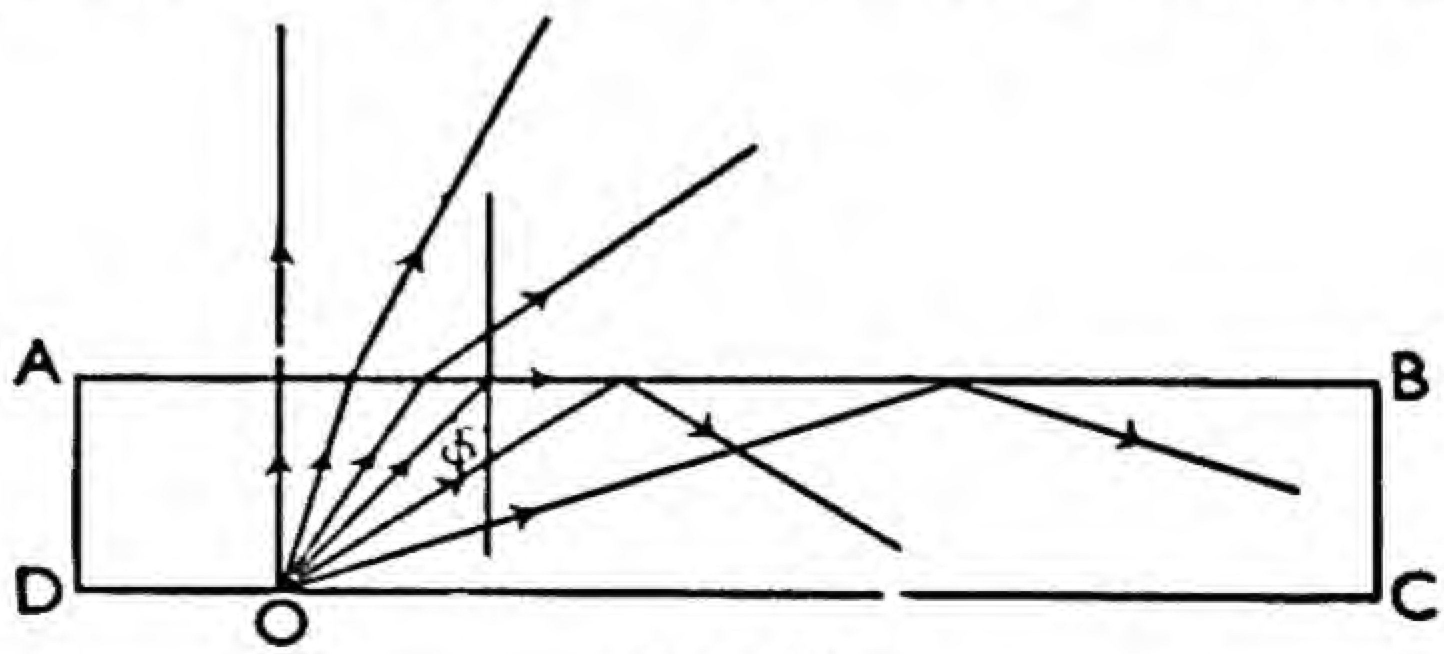


FIG. 31
Rays from object O incident on surface AB separating optically dense medium ABCD from rarer medium above. ϕ Critical Angle

denser than air, the rays are bent away from the normal according to Snell's law, so that if ϕ and θ are the angles of incidence and refraction respectively, the refractive index, μ, of the medium is given by $\mu = \frac{\sin \theta}{\sin \phi}$; or

$$\sin \phi = \frac{\sin \theta}{\mu} \quad . \quad . \quad . \quad . \quad . \quad (3{\cdot}8)$$

where μ, of course, is greater than 1. Now, as ϕ increases, we shall come to a value for which

$$\sin \phi = \frac{1}{\mu} \quad . \quad . \quad . \quad . \quad . \quad (3{\cdot}9)$$

and in that case $\sin \theta = 1$, *i.e.* the emergent ray will travel along the surface of the medium. This value of ϕ is called the *critical angle.* It may be defined as the angle whose sine is the reciprocal of the refractive index of the medium.

What will happen to a ray for which ϕ is greater than the critical angle? Equation (3·8) would give a value of $\sin \theta$ greater than 1, and there is no angle satisfying this condition. We can only make an experiment to see what will happen, and experiment tells us that such a ray is reflected back into the medium. Of course, a small amount of light is reflected back for all angles of incidence, but when the angle passes the critical value, *all* the light is reflected. This phenomenon is known as *total internal reflection.* Obviously it can occur only when the light is travelling from an optically denser towards an optically rarer medium.

Reflecting Prisms: Total internal reflection can be used to give us practically perfect mirrors without any silvering. Fig. 32 shows two prisms of glass, one with angles of 90° and 45°, and the other with angles of 60°, which can be used in this way. It will be noticed that whenever the light crosses a boundary it does so normally, so that there is no refraction. The condition that total reflection shall occur is, of course, that in the first case 45°, and in the second

case 60°, shall be greater than the critical angle, and this is so for ordinary glass.

The same property is useful in changing an inverted

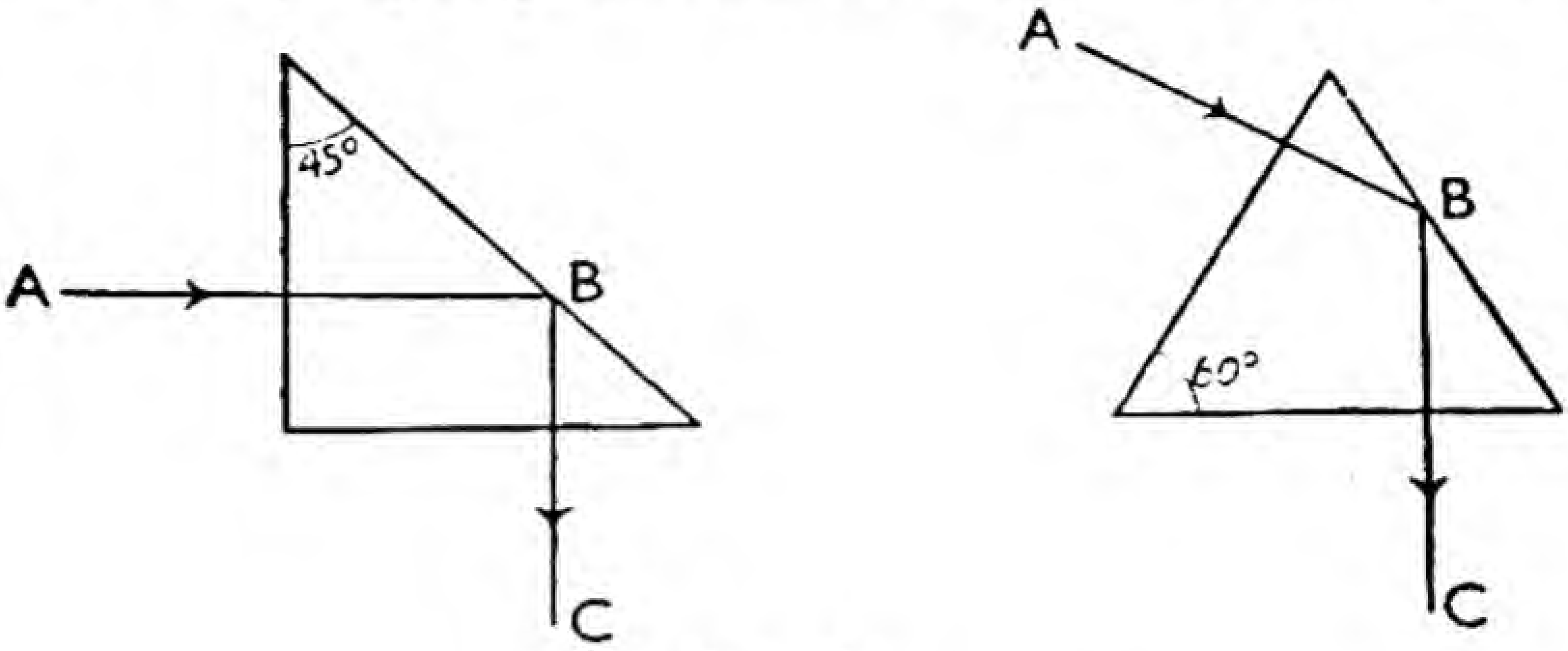


FIG. 32
Totally Reflecting Prisms
ABC Path of Light

image into an erect one. In Fig. 33, pencils A and B, converging to form images of the two ends of an object, are allowed to fall on a 45° isosceles prism, as a result of which

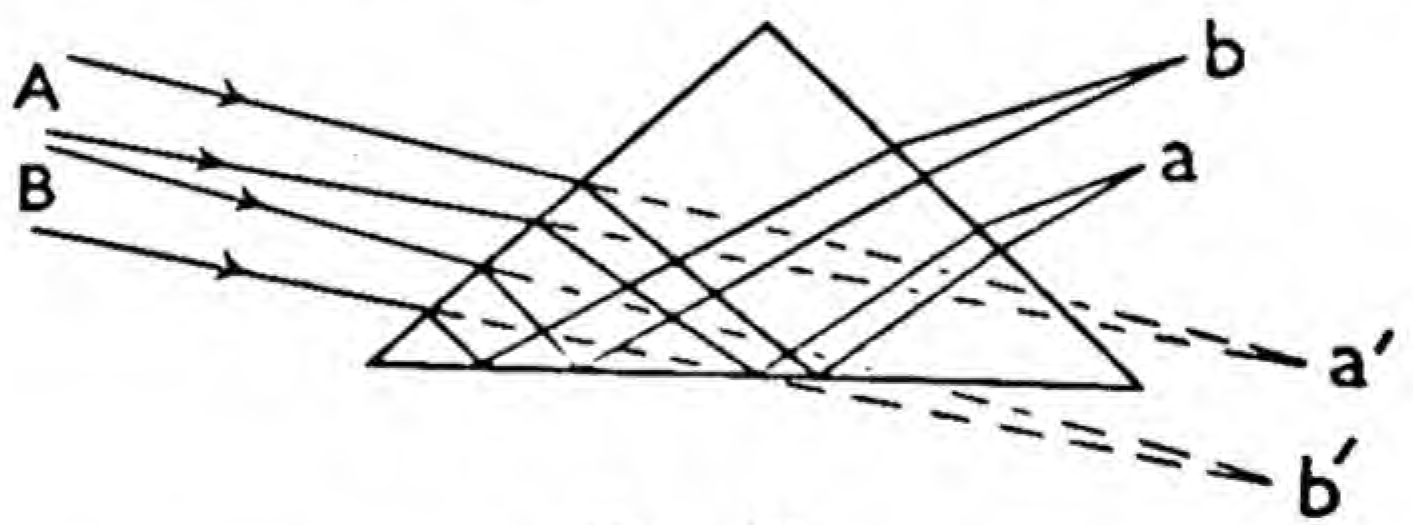


FIG. 33
Inversion of Image by Reflecting Prism
a′b′ Position of Image without Prism
ba Inverted Image produced with Prism

the images are formed at *a* and *b* instead of *a′* and *b′*. It will be noticed that *a′b′* is inverted with respect to *ab*.

Refraction by a Prism

We have seen that if a ray of light passes from air into a succession of parallel-sided blocks of transparent material

and then into the air again, its final direction is the same as its original one. The ray may be displaced laterally, but its direction remains unchanged. We may, however, change the direction of a ray by passing it through a block whose sides are not parallel. The most convenient form of such a block is a triangular prism, represented in Fig. 34. A ray of light incident on the face ABED is bent in the prism towards the base BEFC, and on emerging from the face ADFC is bent again towards the same direction. The edge AD, opposite to the base, is called the *refracting edge* of the prism because, as we shall see, the angle BAC between the faces meeting at this edge

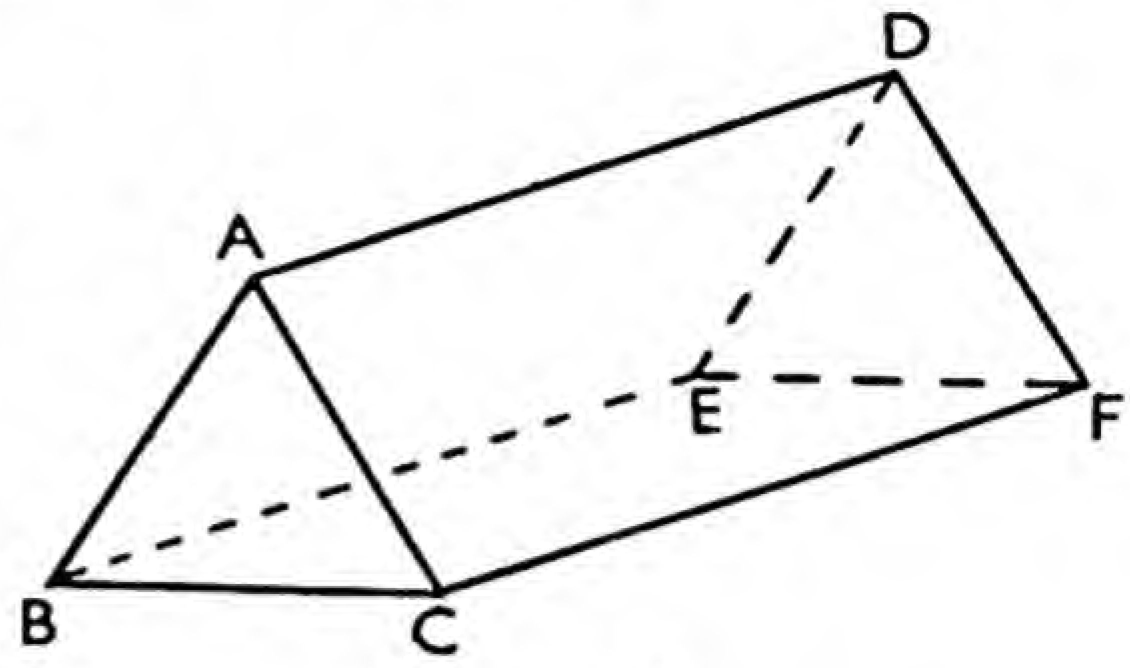


FIG. 34
Triangular Glass Prism
AD Refracting Edge
BAC Refracting Angle
BCFE Base

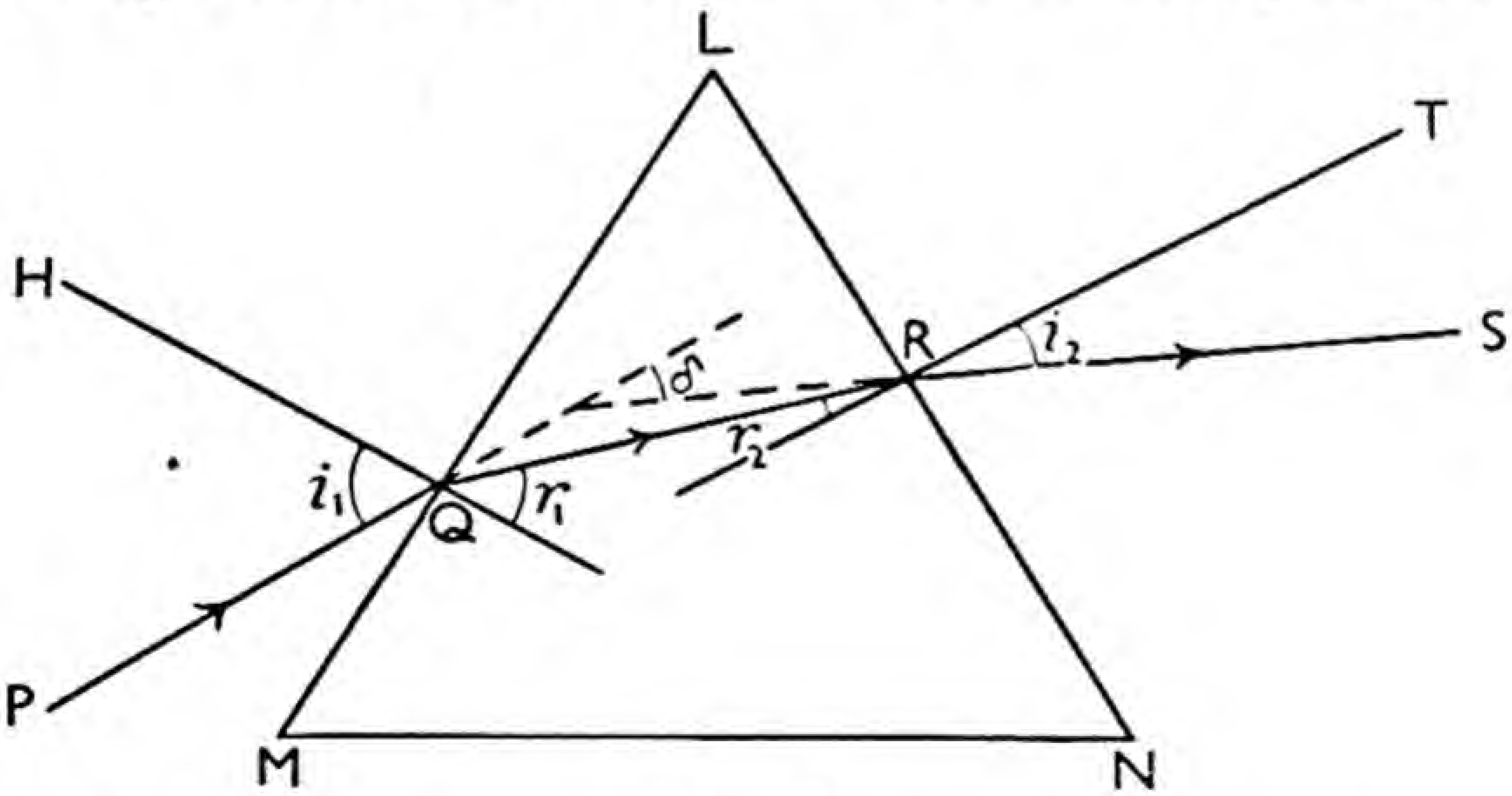


FIG. 35
Passage of Ray through optically dense Prism
PQRS Path of Ray δ Deviation

(known as the *refracting angle*) plays a large part in determining the course of the rays.

We can trace the rays better by considering a section of the prism by a plane perpendicular to the refracting edge. Such a section, LMN, is shown in Fig. 35. Let PQ be an incident ray, making an angle of incidence i_1 with the normal QH. The prism (made of glass, say) being optically denser than air, the ray is bent towards the normal along QR, making angles r_1 ($<i_1$) with QH and r_2 with the normal RT to the face LN. On emerging into the air (an optically rarer medium) the ray is bent away from the normal along RS, making an angle i_2 ($>r_2$) with RT. The net result of the passage, then, is that the direction of the ray is changed from PQ to RS. The angle δ between these directions is called the *deviation* produced by the prism.

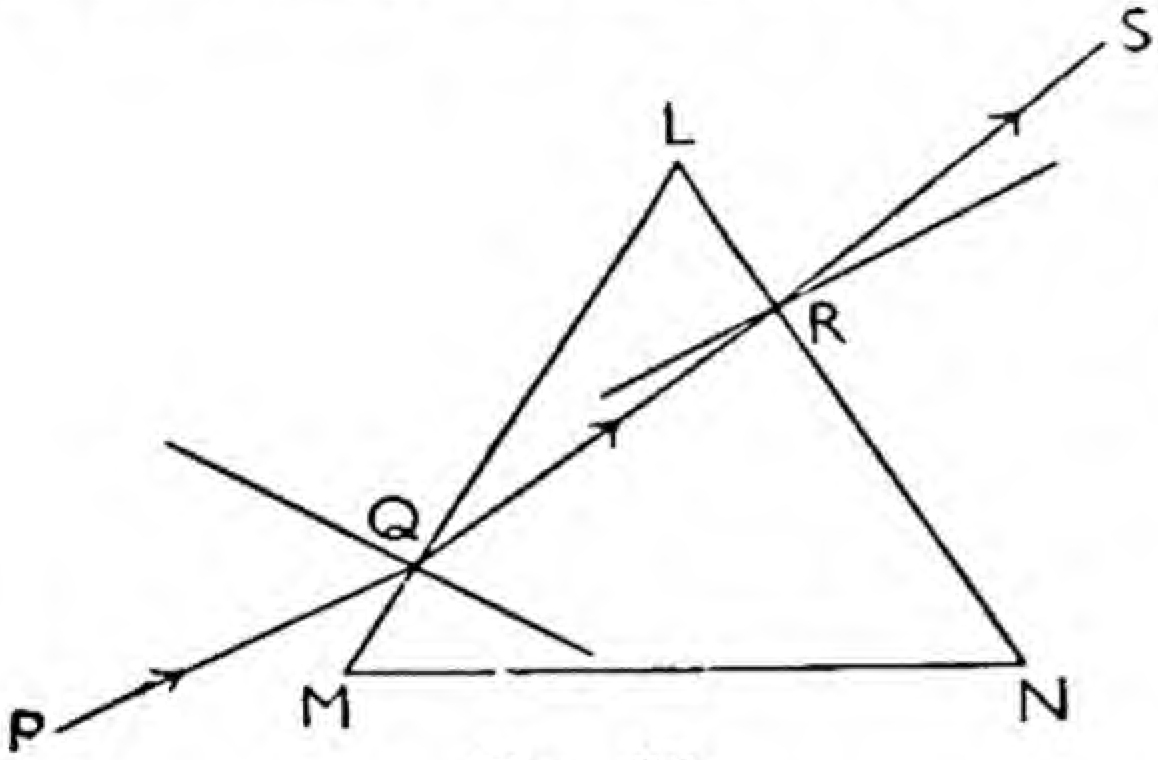


Fig. 36
Passage of Ray PQRS through optically rare Prism

It may be noticed in passing that if the prism is of rarer material than its surroundings — a hollow glass prism, say, in a large vessel of water—the deviation is away from the base of the prism instead of towards it, as Fig. 36 shows.

Reverting to the ordinary case of a dense prism, we may notice that a narrow pencil of light forms an image, just as with a parallel-sided block, and, as in that case, the position of the image varies with the particular pencil chosen. Hence, if we are looking directly at an object O, and a glass prism is then interposed in the path, we cease to see the object, and have to look in a quite different direction in order to see an image of it. This is shown in Fig. 37, from which it is clear that instead of placing the eye at E_1, we must now place it at some position such as E_2 or E_3 and look in a different direction.

Angle of Minimum Deviation: Let us observe a particular pencil, *e.g.* that forming the image I_2 in Fig. 37, and slowly rotate the prism about its centre. Then the angle of incidence of this pencil on the face LM will gradually change. The directions of the pencil in the prism and on emergence into the air again will therefore also change, by Snell's law. Consequently the deviation will change. Now it is found

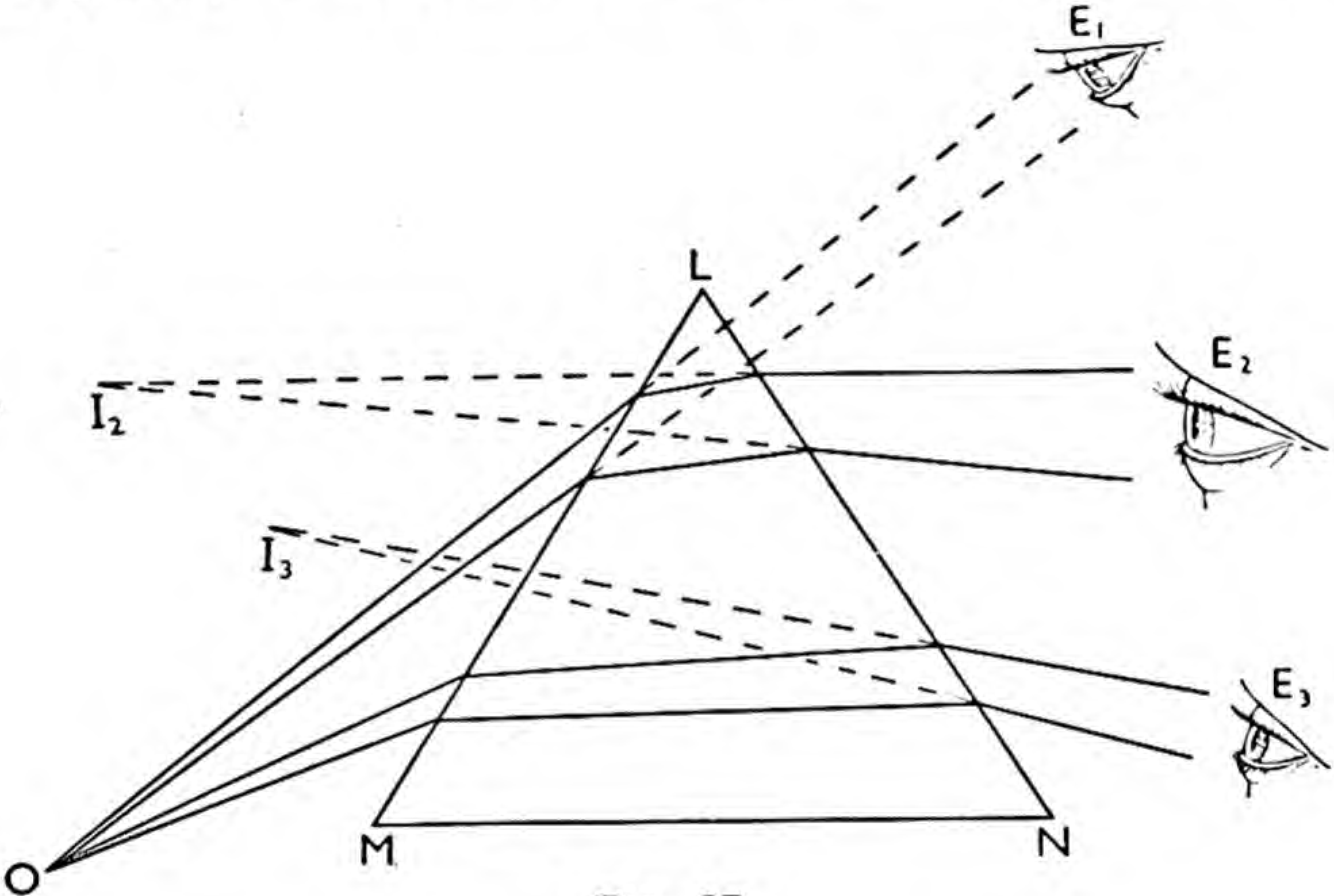


FIG. 37

Images formed by Refraction through Prism

O Object I_2, I_3 Images formed by different pencils of light

that there is a certain position of the prism, *i.e.* a certain angle of incidence, for which the deviation is a minimum; if the angle of incidence is either slightly greater or slightly less than this, the deviation is greater.

This can be proved mathematically, from the law of refraction, but we shall take it here as established by experiment. We can then prove that when the deviation is a minimum, the light passes symmetrically through the prism, *i.e.* the path of the ray inside the prism forms an isosceles

triangle with the faces LM and LN. For in Fig. 38 let ABCD be the symmetrical ray, and let us suppose that it is *not* the ray for which the deviation is a minimum, but that some other ray A′BC′D′ has this property. Then it is clear that a ray D″CB′A″, entering the prism from the other side and situated symmetrically, with regard to ABCD, to the ray A′BC′D′, will have the same deviation, for its angles of incidence and refraction will have the same values. Also, since a ray refracted back along its own path keeps to that path throughout, this ray, sent in the direction A″B′CD″,

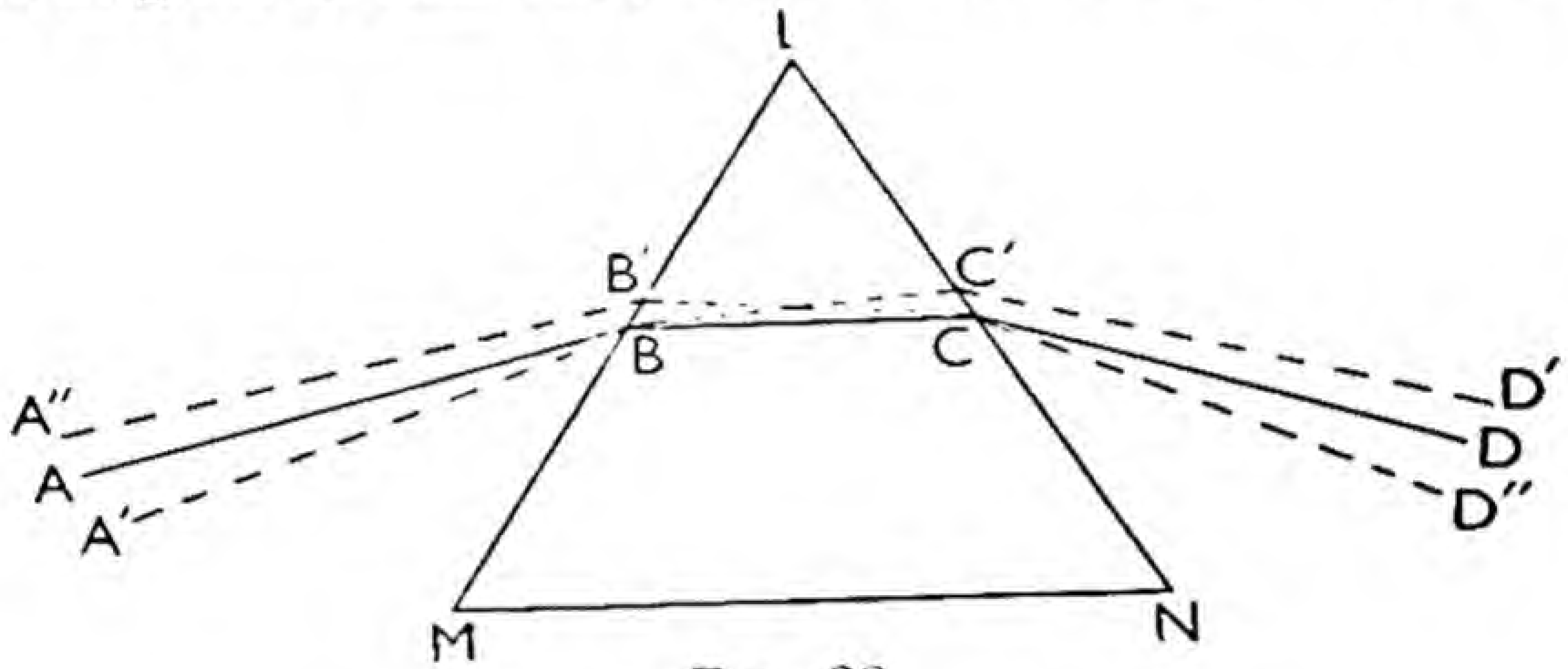


FIG. 38
Paths of Rays through Prism
ABCD Symmetrical Path of Minimum Deviation

will also have the minimum deviation. We have then two rays, A′BC′D′ and A″B′CD″, both having the minimum deviation. But we know from experiment that there is only one ray having this property. Hence this cannot be a ray different from the symmetrical one.

Measurement of Refractive Index

We can express the refractive index of the material of a prism in terms of the angle of minimum deviation and the refracting angle of the prism. For, in Fig. 39, in which the lettering is the same as in Fig. 35, and the refracting angle is called α, we see that the exterior angle δ (= VWR) is equal to WQR + QRW, *i.e.*

$$\delta = i_1 - r_1 + i_2 - r_2. \qquad (3{\cdot}10)$$

Now in the quadrilateral LQXR the angles at Q and R are right angles. Hence the angles at L and X together

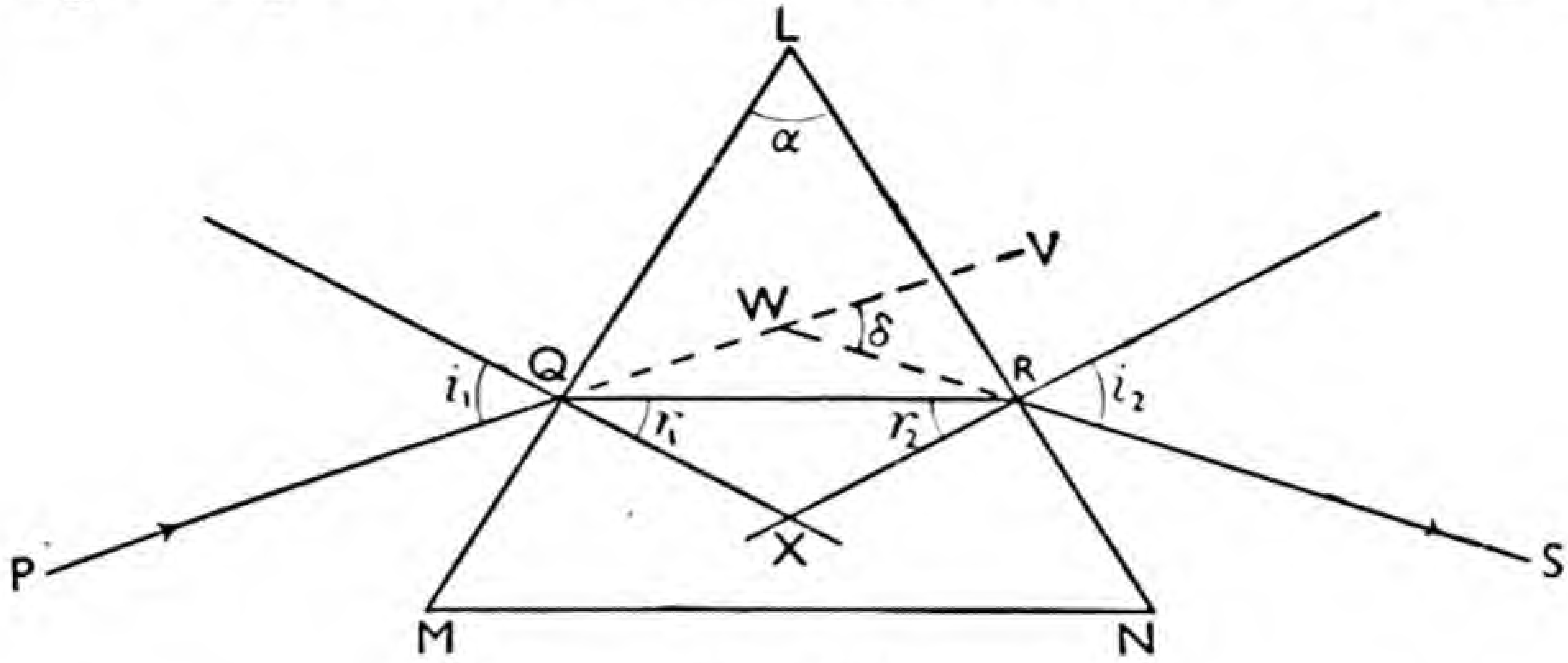


Fig. 39

Deviation δ of Ray PQRS passing symmetrically through Prism

make two right angles, so that $\alpha = \pi - X$. But in the triangle QXR the angles $r_1 + r_2$ are equal to $\pi - X$. Hence

$$\alpha = r_1 + r_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (3 \cdot 11)$$

so that, from (3·10),

$$\delta = i_1 + i_2 - \alpha. \quad . \quad . \quad . \quad . \quad . \quad (3 \cdot 12)$$

Now, when the deviation is a minimum, $i_1 = i_2$ and $r_1 = r_2$, since the ray passes symmetrically through the prism. Calling the common values i and r, we then have, from (3·11) and (3·12),

$$\alpha = 2r \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3 \cdot 13)$$

$$\text{and} \quad \delta = 2i - \alpha \quad . \quad . \quad . \quad . \quad . \quad . \quad (3 \cdot 14)$$

Hence

$$\mu = \frac{\sin i}{\sin r} = \frac{\sin \dfrac{\alpha + \delta}{2}}{\sin \dfrac{\alpha}{2}} \quad . \quad . \quad . \quad . \quad . \quad (3 \cdot 15)$$

This provides a very accurate method of measuring the refractive index of a material.

Refraction at Curved Surfaces

Snell's law applies to a curved as well as to a plane surface, and we can show that refraction at such a surface produces an image under the same conditions as those controlling the formation of images by curved mirrors. If surfaces of large aperture are involved, then the images lie on a caustic curve, as before ; we shall not consider these cases.

Concave Refracting Surface

Let us take a spherical concave surface, separating two media whose refractive indices with respect to air are respectively

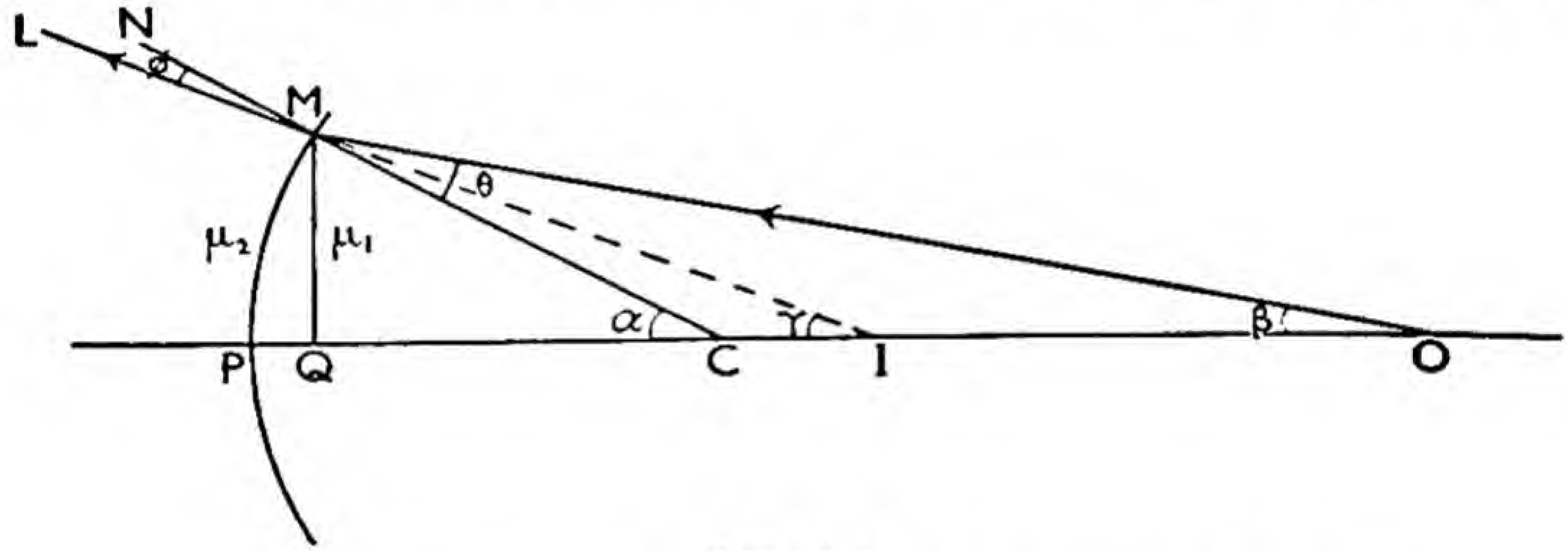


Fig. 40

Refraction of ray OML at Concave Spherical Surface

O Object I Virtual Image

μ_1 and μ_2 ($\mu_1 < \mu_2$), and suppose the object is an axial point source of light in the first medium, so that the light travels from the rarer towards the denser medium. The definitions of pole, axis, etc., are the same as for mirrors, and we use the same symbols and sign convention for the positions of the object, image, and centre of curvature.

A ray from O through C meets the surface of separation normally, and goes straight through (Fig. 40). Another ray, meeting the mirror at M, is bent towards the normal along ML, and obviously diverges from the axis wherever on the

axis O may be, so that no real image is formed. It proceeds, however, as though it came from a point I, lying between O and C, and this is the position of the virtual image formed.

We shall now derive a relation connecting the positions of the object and image with the refractive indices and the radius of curvature, r, from which we shall see that all rays from O (the aperture of the surface being small) proceed on reflection as though they came from a point very close to I, so that there is actually an image of the object at I. Let the angle of incidence OMC be θ, and the angle of refraction NML = CMI be ϕ. Then, from equation (3·7),

$$\frac{\sin\theta}{\sin\phi} = \frac{\mu_2}{\mu_1} \quad . \quad . \quad . \quad . \quad . \quad (3 \cdot 16)$$

or, since the angles must be small for surfaces of small aperture,

$$\frac{\theta}{\phi} = \frac{\mu_2}{\mu_1} \quad . \quad . \quad . \quad . \quad . \quad (3 \cdot 17)$$

approximately. Let the angles MCP, MOP, and MIP be α, β, and γ respectively. These angles also will be small, and their sines will bear the same ratios to one another as the angles themselves.

Now, from Fig. 40,

$$\alpha = \gamma + \phi = \beta + \theta \quad . \quad . \quad . \quad . \quad . \quad (3 \cdot 18)$$

$$\text{Also} \quad \left.\begin{aligned} \frac{\text{MQ}}{r} &= \sin\alpha \\ \frac{\text{MQ}}{\text{MO}} &= \sin\beta = \sin(\alpha - \theta) \\ \frac{\text{MQ}}{\text{MI}} &= \sin\gamma = \sin(\alpha - \phi) \end{aligned}\right\} \quad . \quad . \quad . \quad (3 \cdot 19)$$

Hence

$$r\sin\alpha = \text{MO}\sin(\alpha - \theta) = \text{MI}\sin(\alpha - \phi) \quad . \quad (3 \cdot 20)$$

Also, since M is near P we may take MO and MI as equal to u and v respectively, and, substituting angles for their sines, we have

$$r\alpha = u(\alpha - \theta) = v(\alpha - \phi) \quad . \quad . \quad . \quad (3{\cdot}21)$$

$$\text{whence} \quad \alpha = \frac{u\theta}{u - r} = \frac{v\phi}{v - r} \quad . \quad . \quad . \quad (3{\cdot}22)$$

$$\text{We have, therefore,} \quad \frac{\theta}{\phi} = \frac{\mu_2}{\mu_1} = \frac{v(u - r)}{u(v - r)} \quad . \quad . \quad . \quad (3{\cdot}23)$$

$$\text{which reduces to} \quad \frac{\mu_2}{v} - \frac{\mu_1}{u} = \frac{\mu_2 - \mu_1}{r} \quad . \quad . \quad . \quad (3{\cdot}24)$$

This is the equation we are seeking. It will be noticed that β does not appear in it, so that the value of v for a given value of u (*i.e.* the position of I for a given position of the object) is the same for all rays near the axis.

If the first medium is air, then $\mu_1 = 1$, and the equation becomes

$$\frac{\mu}{v} - \frac{1}{u} = \frac{\mu - 1}{r} \quad . \quad . \quad . \quad . \quad (3{\cdot}25)$$

where the refractive index of the second medium is called simply μ.

Equation (3·24) may be written in the form

$$\frac{1}{v} = \frac{1}{\mu_2}\left[\frac{\mu_1}{u} + \frac{\mu_2 - \mu_1}{r}\right] . \quad . \quad . \quad . \quad (3{\cdot}26)$$

From this it is clear that a real image cannot be formed when $\mu_2 > \mu_1$, for with a concave surface u and r are necessarily positive, so that v must be positive also. If, however, $\mu_2 < \mu_1$, *i.e.* the light travels from a denser towards a rarer medium, v may become positive or negative according to the resultant sign of the quantity in square brackets. The

two cases are illustrated in Fig. 41, where (*a*) shows a real image and (*b*) a virtual image.

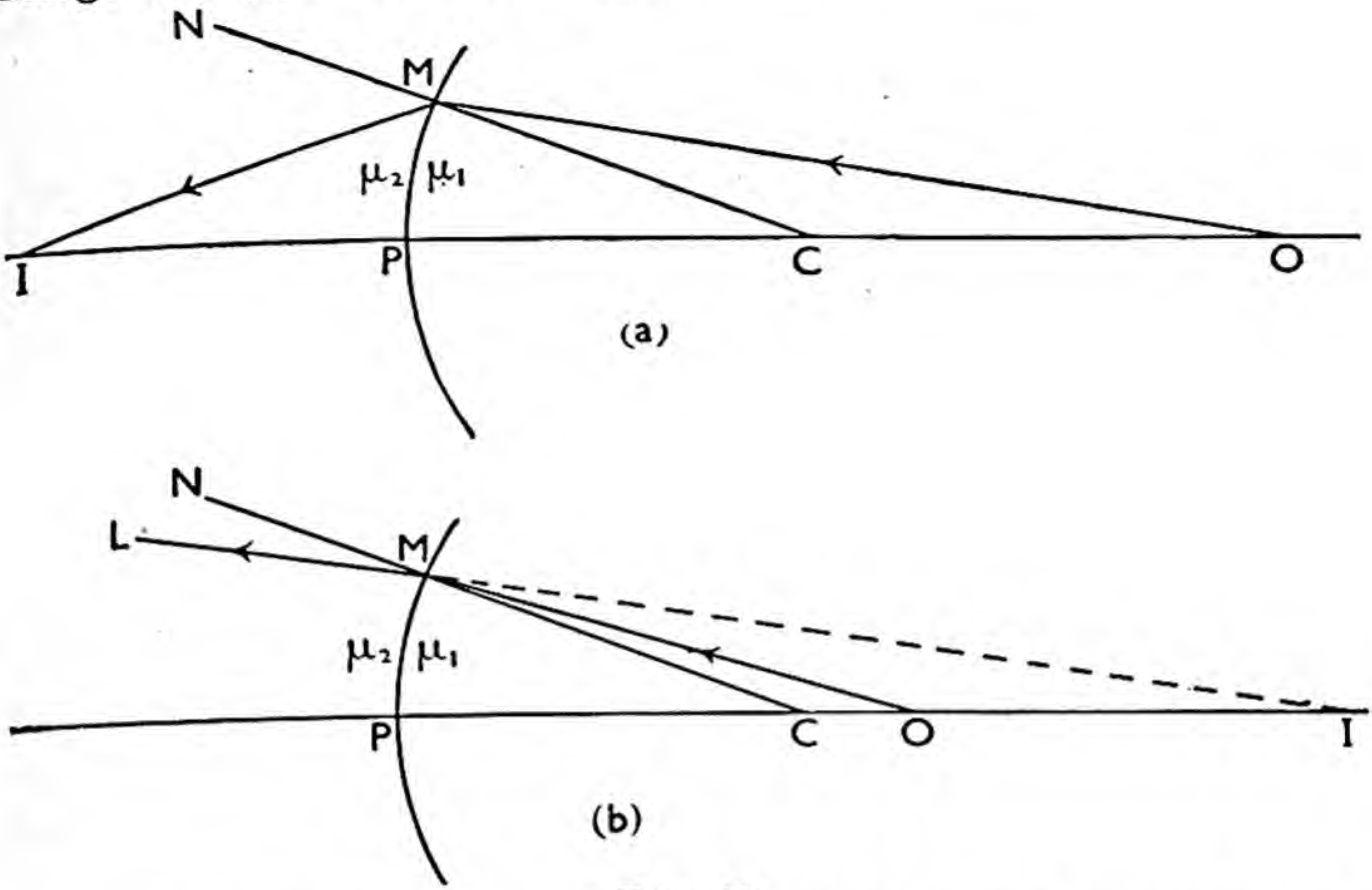


FIG. 41
Refraction of ray OM on crossing Concave Spherical Surface when travelling from denser to rarer medium
O Object I (*a*) Real Image (*b*) Virtual Image

Convex Surfaces

Refraction at a convex surface can be examined in the same way. For want of space we give merely the results. The same formula (3·24) is found to hold, but it must be remembered that here r is negative. The image obtained may be either real or virtual if $\mu_2 > \mu_1$, but is always virtual if $\mu_2 < \mu_1$.

Objects of Finite Size

A small axial object of finite size forms with both concave and convex surfaces an axial image, whose size, I, relative to that of the object, O, can easily be shown to be given by

$$\frac{\mathrm{I}}{\mathrm{O}} = \frac{v\mu_1}{u\mu_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3\cdot27)$$

A real image is always inverted, and a virtual image erect. This is indicated by the sign of the expression in (3·27).

Non-interchangeability of Object and Image

One important point should be noticed. The positions of object and image are not interchangeable as they are with images formed by reflection ; that is to say, the image of an object placed at I is not situated at O. This follows from formula (3·24), which does not remain the same when u and v are interchanged, and the student may easily verify it graphically.

Lenses

The most important application of these results is to the formation of images by lenses. A lens consists of a piece of transparent material (usually glass) with one or both surfaces curved. We shall consider only cases in which the curved surfaces are spherical and of small curvature and small aperture, so that the lenses are very thin.

Types of Lenses

When light from an object diverges towards a lens, the effect of the lens is to make the light diverge either more or less, or to make it converge. If the divergence is increased, the lens is called a *diverging* lens ; if it is decreased or changed to convergence, the lens is called a *converging* lens. When the material of the lens is denser than that of its surroundings (usually air), a diverging lens is thinner, and a converging lens thicker, at the centre than at the edges. When the material is rarer than its surroundings these relations are reversed. Fig. 42 shows a few examples of each type of lens.

A line bisecting a lens normally is known as the *axis* of the lens ; AB is the axis of the lenses in Fig. 42. The centre of the portion of the axis within the lens is known as the *optical centre* of the lens. We shall consider only lenses so thin that the whole of the axis within the lens may be taken as the optical

centre, its length being negligible compared with the distances u and v of the object and image.

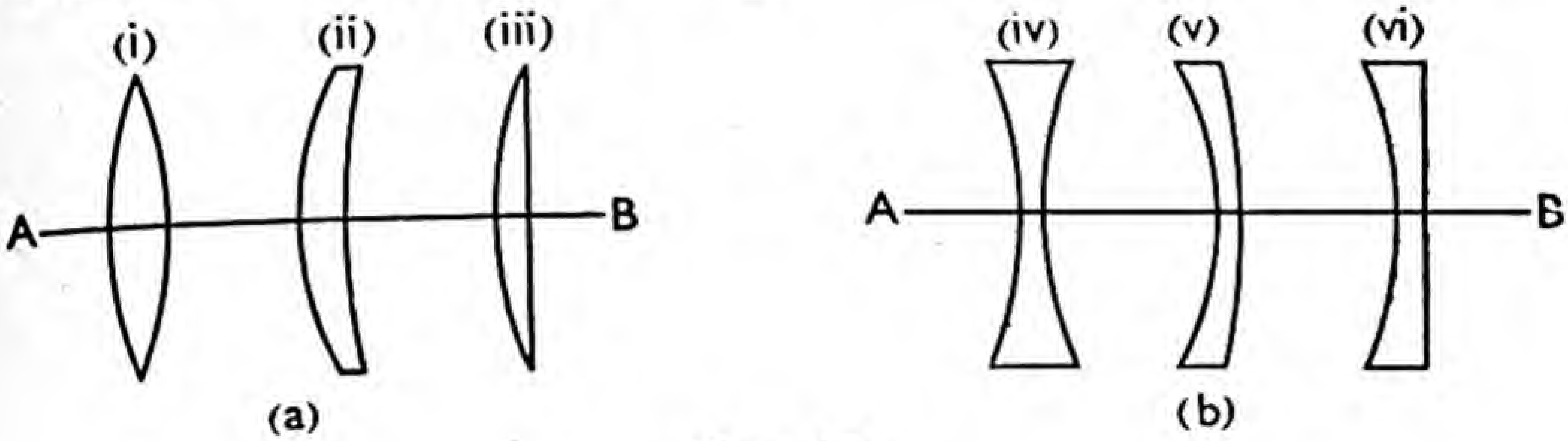


FIG. 42

(*a*)
Converging Lenses
(i) Double Convex
(ii) Converging Meniscus
(iii) Plano-convex

(*b*)
Diverging Lenses
(iv) Double Concave
(v) Diverging Meniscus
(vi) Plano-concave

Images formed by a Lens

We may easily show that a lens produces an image. We have just seen that light from an axial point is refracted at the first surface, and proceeds as though it came from a point, I_1, say. It falls on the second surface, therefore, as though it came from an object at I_1, and thus forms an image of that supposed object. This final image is the image of the object which the lens produces.

Let us consider a " double convex " lens ((i) in Fig. 42), with radii of curvature r_1 and r_2, as an example : the formation of images by other lenses can be dealt with similarly. Let μ (> 1) be the refractive index of the material (glass, say), and let the lens be situated in air. An object at O, distant u from the lens (Fig. 43), would produce, by refraction at the first surface alone, an image at I_1 at a distance v_1 given by

$$\frac{\mu}{v_1} - \frac{1}{u} = \frac{\mu - 1}{r_1} \quad . \quad . \quad . \quad . \quad . \quad (3\cdot28)$$

This image acts as an object for the second surface, and the light crossing that surface travels from glass into air, so in

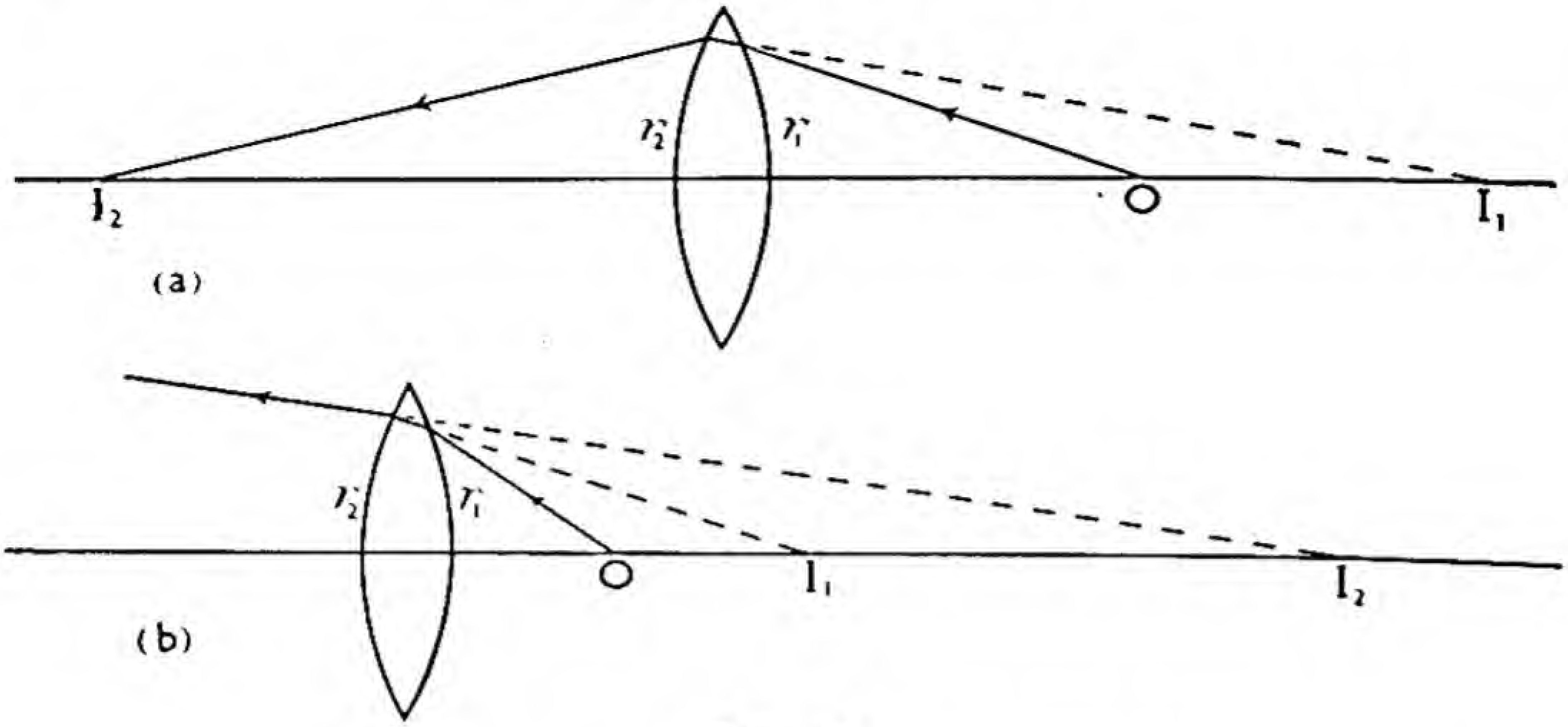


FIG. 43
Images I_2 formed by Double Convex Lens
(*a*) Real Image (*b*) Virtual Image

applying equation (3·24) we must put $\mu_2 = 1$, $\mu_1 = \mu$, and $u = v_1$. Hence the position, v, of the final image I_2 is given by

$$\frac{1}{v} - \frac{\mu}{v_1} = \frac{1 - \mu}{r_2} \quad . \quad . \quad . \quad . \quad . \quad (3\cdot29)$$

Adding equations (3·28) and (3·29) we thus obtain

$$\frac{1}{v} - \frac{1}{u} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad . \quad . \quad . \quad (3\cdot30)$$

It must be remembered that r_1 is negative and r_2 positive, so that the right-hand side does not vanish when r_1 and r_2 are numerically equal, as they often are.

The equation may be written

$$\frac{1}{v} = \frac{1}{u} + (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad . \quad . \quad . \quad (3\cdot31)$$

From the signs of r_1 and r_2 we see that the second term is necessarily negative, but u must be positive, so that v may be positive or negative, according to the relative magnitudes

of the quantities involved. In the former case we have a virtual image, and in the latter a real image. The two cases are illustrated in Fig. 43 (*b*) and (*a*) respectively.

Formula (3·30) holds for all types of lens when the symbols are given the proper signs according to the rule.

Principal Focus and Focal Length

It will be noticed that the right-hand side of equation (3·30) is constant for a given lens. We may easily show that it is

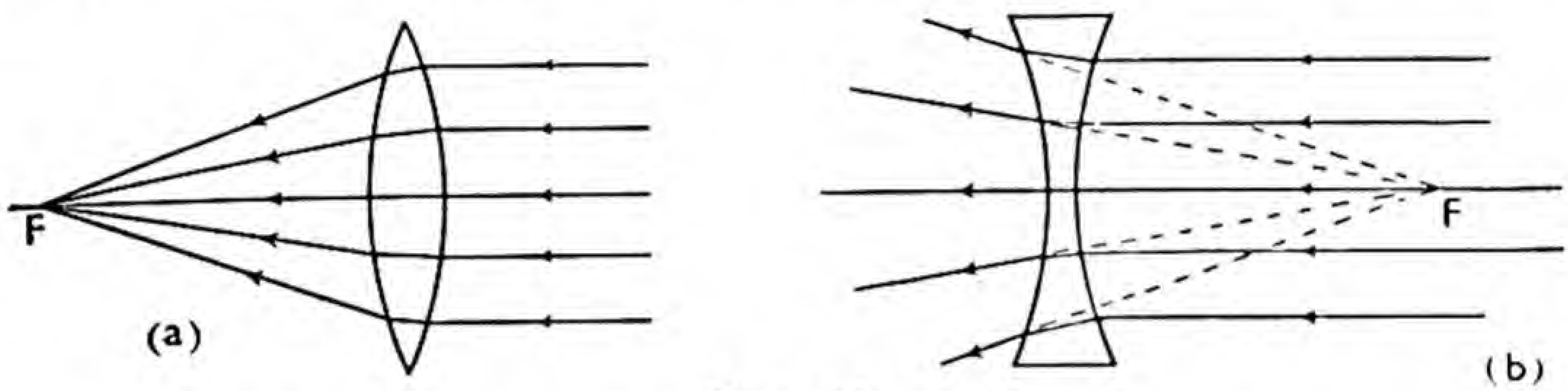


FIG. 44

Formation of Image at Principal Focus F of Lens by Rays parallel to Axis
(*a*) Converging Lens—Real Image (*b*) Diverging Lens—Virtual Image

the reciprocal of the distance from the lens of the image of an object at infinity. For, putting $u = \infty$, we obtain

$$\frac{1}{v} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad . \quad . \quad . \quad . \quad (3\cdot32)$$

This constant quantity is usually written $\frac{1}{f}$, and f is called the *focal length* of the lens, while the point F at which the image is formed is called the *principal focus*. With a converging lens a real image, and with a diverging lens a virtual image, is formed by parallel rays, as Fig. 44 shows.

We may now write equation (3·30) as

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f} \quad . \quad . \quad . \quad . \quad . \quad (3\cdot33)$$

and f is clearly negative for a converging lens and positive for a diverging lens.

Type and Position of Image

In order to see what type and position of image will result from a given position of the object, we make use of the fact, which will be fairly obvious, that when v is negative we have a real image, and when it is positive we have a virtual

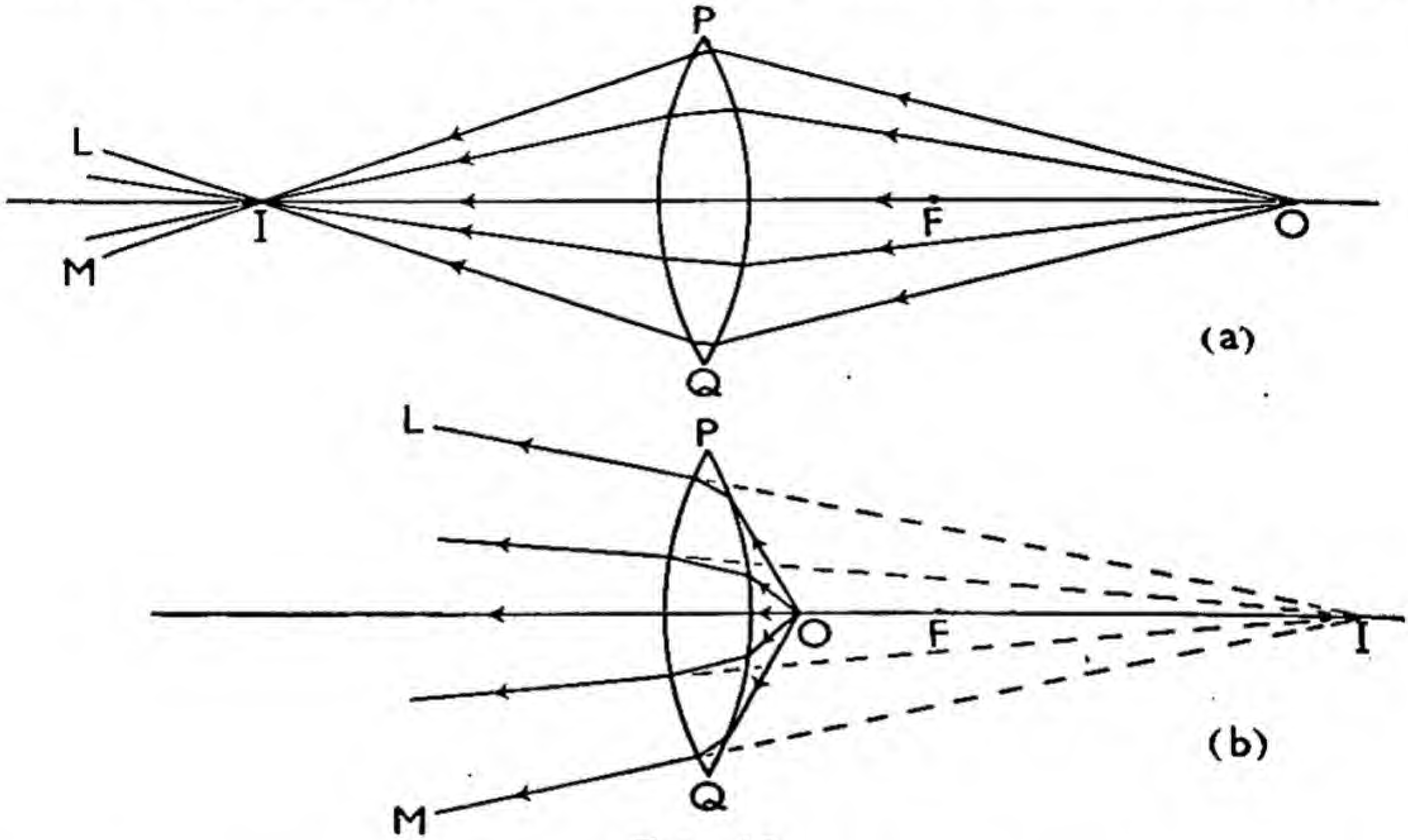


FIG. 45

(a) Real and (b) Virtual Image produced by Object (a) outside and (b) inside the Principal Focus of a Converging Lens

image. Consider first a converging lens. We then have, from (3·33),

$$\frac{1}{v} = \frac{1}{u} + \frac{1}{f} \quad . \quad . \quad . \quad . \quad . \quad (3{\cdot}34)$$

where u is positive and f negative. Hence v will be negative (*i.e.* the image will be real) only so long as $u > f$ numerically. That is, an object outside the principal focus forms a real image on the other side of the lens, while an object inside the principal focus forms a virtual image on the same side of the lens. These cases are, respectively, (a) and (b) of Fig. 43, and they are shown in Fig. 45 with more detail. In order to

see the image the eye must be placed within the cone of light LIM in either case, unless the real image is received on a screen which scatters its light as previously explained.

For a diverging lens f is positive, and, u also being positive, v must be positive and smaller than u (see (3·34)) for all positions of the object. The image is, therefore, invariably a virtual one, situated between the object and the lens (Fig. 46).

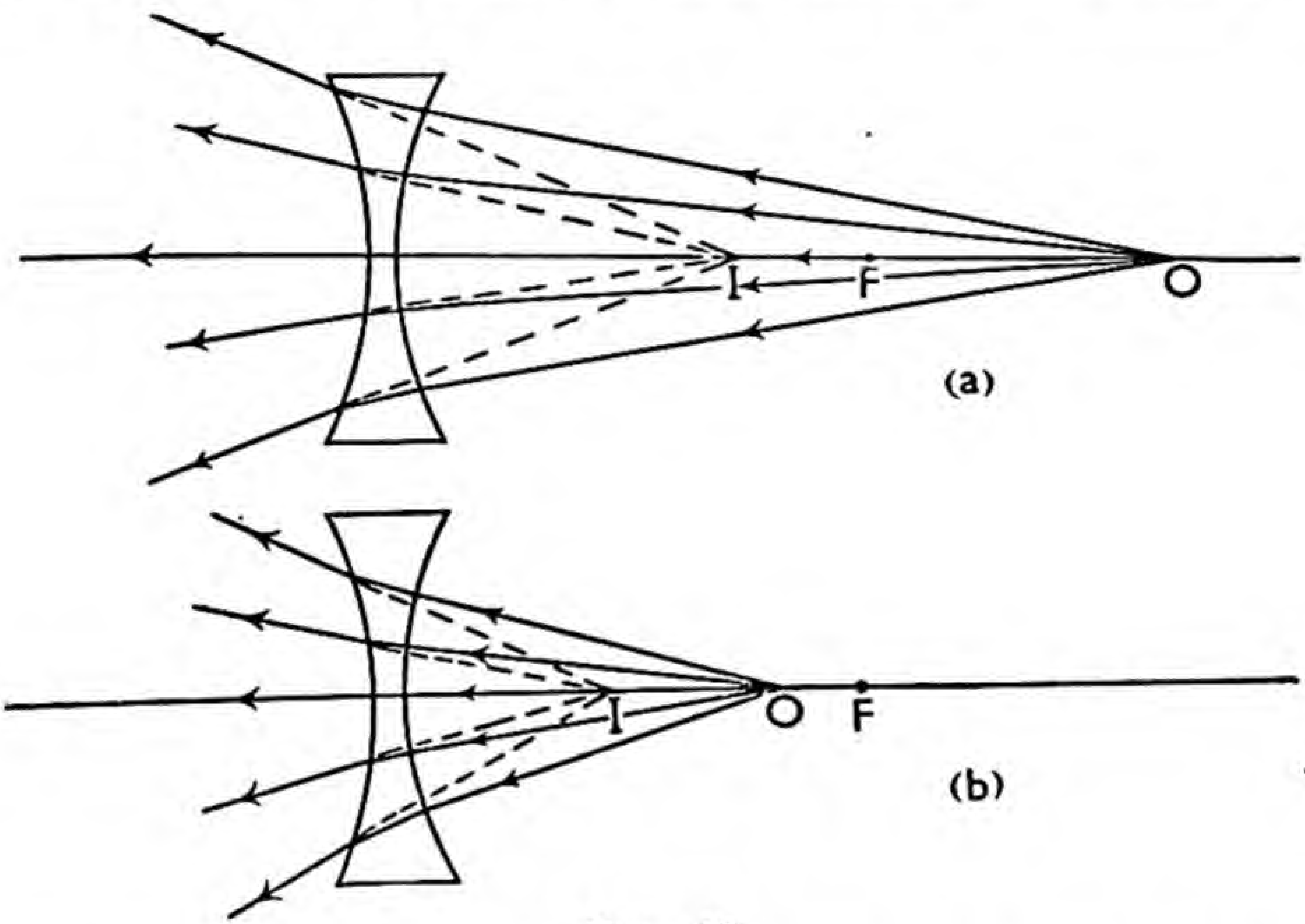


FIG. 46

Virtual Image formed by Diverging Lens of Object (*a*) outside and (*b*) inside Principal Focus

Extended Objects: To find the position and form of the image of an extended object near the axis we proceed as follows. From each point of the object (for example, the point A in Fig. 47) draw two rays, one parallel to the axis and the other towards the optical centre of the lens. The former, after passage through the lens, proceeds towards the principal focus on the far side (or as if from the principal focus on the near side if the lens is a diverging one), while the other, falling on what is effectively a parallel-sided piece

of glass since the surfaces are parallel at the centre of the lens, goes straight through, with a very slight lateral deviation which is negligible when the lens is thin. The point A′ at

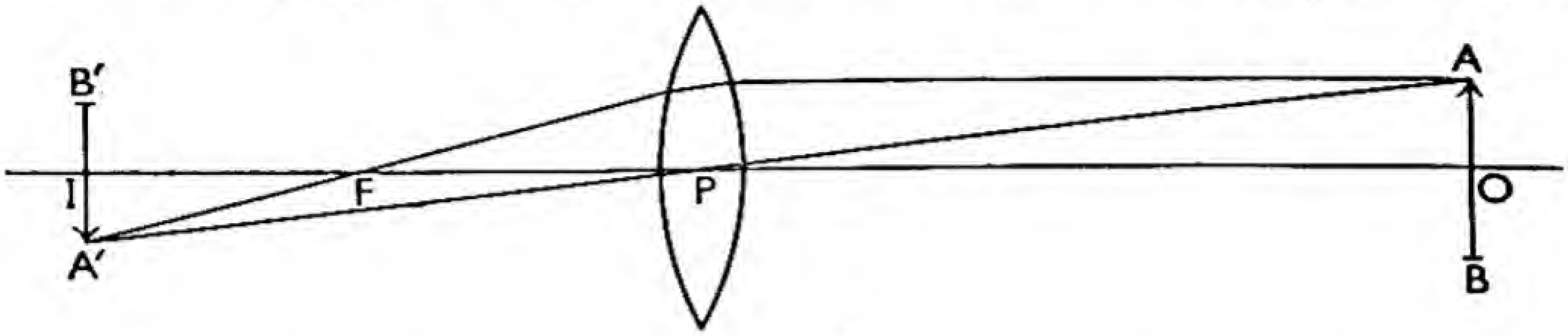


FIG. 47
Formation of Real Inverted Image B′A′ of Extended Axial Object AB by Converging Lens

which these rays meet is the image of A. In this way the whole image may be built up.

It is clear from Fig. 47 that

$$\frac{OA}{IA'} = \frac{PO}{PI} = \frac{u}{v} \text{ numerically} \quad . \quad . \quad . \quad (3{\cdot}35)$$

Hence, as with spherical mirrors, the sizes of image and object are in the ratio of their distances from the lens. The same rule may be seen to hold for both types of lens and for real and virtual images, and it may easily be verified that a real image is inverted and a virtual image erect.

Combination of Lenses

When a lens of larger or smaller focal length than those available is required, it may often be obtained by placing two lenses together. In Fig. 48 the case of two converging

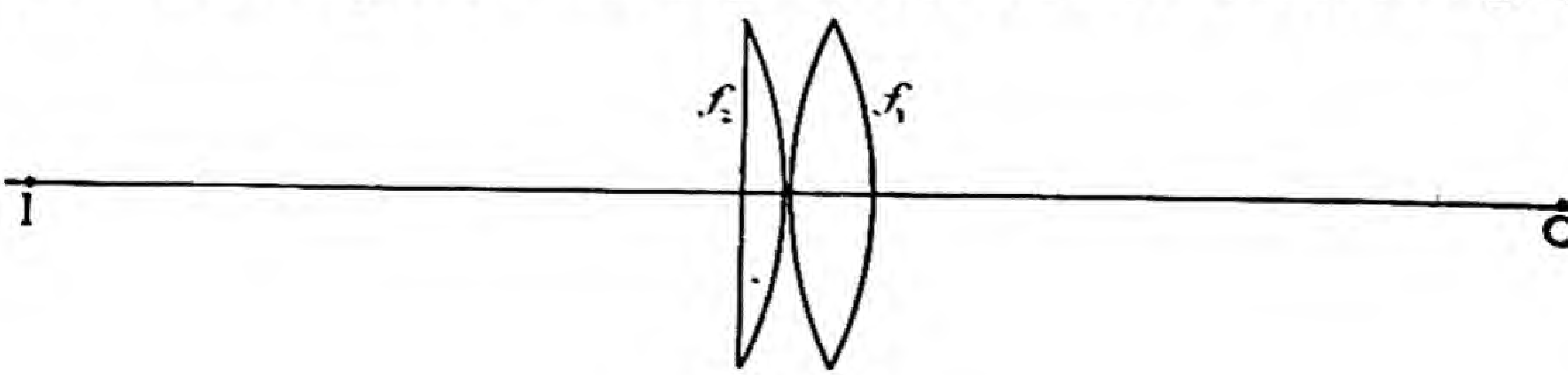


FIG. 48
Combination of Lenses

lenses is illustrated, the focal lengths being f_1 and f_2. To find the position of the image formed by such a combination we proceed in the manner already familiar ; we find first the position of the image which would be formed by the lens on which the light falls first, and treat that image as an object for the second lens. Thus, suppose the image of O which the first lens would form is at I, distant v_1 from the lens. Then

$$\frac{1}{v_1}-\frac{1}{u}=\frac{1}{f_1} \quad . \quad . \quad . \quad . \quad . \quad (3{\cdot}36)$$

For the second lens the distance of the object is v_1, so that if v is the distance of the final image we have

$$\frac{1}{v}-\frac{1}{v_1}=\frac{1}{f_2} \quad . \quad . \quad . \quad . \quad . \quad (3{\cdot}37)$$

Adding these two equations we obtain

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f_1}+\frac{1}{f_2} \quad . \quad . \quad . \quad . \quad . \quad (3{\cdot}38)$$

Comparing this with (3·33) we see that the combination acts as a single lens of focal length f, where

$$\frac{1}{f}=\frac{1}{f_1}+\frac{1}{f_2} \quad . \quad . \quad . \quad . \quad . \quad (3{\cdot}39)$$

If, then, we want a lens of short focus, we choose f_1 and f_2 of the same sign—whichever sign we want. For a long focus lens we choose them of opposite signs.

Spherical Aberration

It may seem that the restrictions we have introduced—considering only thin lenses and rays making small angles with the axis—make our considerations of only theoretical interest, but that is far from being the case. In many of the uses to which lenses are put we keep within these restrictions, and our formulae can be applied without appreciable error. Occasion-

ally, however, we have to take into account the thickness of the lens, and deal with wide-angle rays. This may impose serious problems on the lens designer, and special forms have to be computed for particular problems. The effect of the peripheral rays in distorting an image is known as *spherical aberration*, as with mirrors, and often great ingenuity is required to minimize it.

The simplest way of reducing the spherical aberration of a lens is to "stop" it, *i.e.* cover it with a screen having a small central hole, so that only the central parts receive the light. This, however, gives a fainter image, since light is excluded, but that is sometimes less important than securing the best definition in the image.

EXERCISES

1. Explain with a diagram why a straight object appears bent when partly immersed in water.

2. State Snell's law, and prove that when a beam of light passes through a succession of media of different refractive indices, its final direction depends only on the initial direction and the refractive index of the last medium.

3. Show how to find the position of the image of a small object seen through a parallel-sided glass block. If the block rests on the object, prove that the ratio of the depths below the upper surface, of the object and its image when viewed normally, is equal to the refractive index of the glass.

4. Explain, with a drawing, the formation of several images of a source of light held in front of a thick mirror silvered on the back surface.

5. Calculate, and verify by geometrical construction, the deviation of rays of light incident on a glass prism of refracting angle 60° and refractive index 1·6, when the angles of incidence are 30°, 40°, 50°, 55°, 60°, 70°, and 80° respectively. Plot a curve showing how the deviation varies with the angle

of incidence, and so determine the angle for minimum deviation. By drawing the path of the corresponding ray, prove that it passes symmetrically through the prism.

6. Explain what is meant by total internal reflection, and state how the phenomenon may be put to practical use.

7. Prove formula (3·24) for refraction at a *convex* surface, and describe how the position and character of the image change as the object is moved along the axis.

8. Find the ratio of the focal lengths of a glass lens in water and in air, if the refractive indices of glass and air are respectively $1\frac{1}{2}$ and $1\frac{1}{3}$.

9. Find an expression for the focal length of a combination of lenses whose focal lengths are f_1 and f_2. Under what conditions is the focal length of the combination (*a*) greater and (*b*) less, numerically, than either f_1 or f_2?

CHAPTER IV

OPTICAL INSTRUMENTS AND THE VELOCITY OF LIGHT

The Eye

In this chapter we consider mainly instruments in which the optical parts consist entirely of lenses. The most indispensable of these is the eye, the chief parts of which are the double-convex *crystalline lens* L (Fig. 49), a transparent horny covering

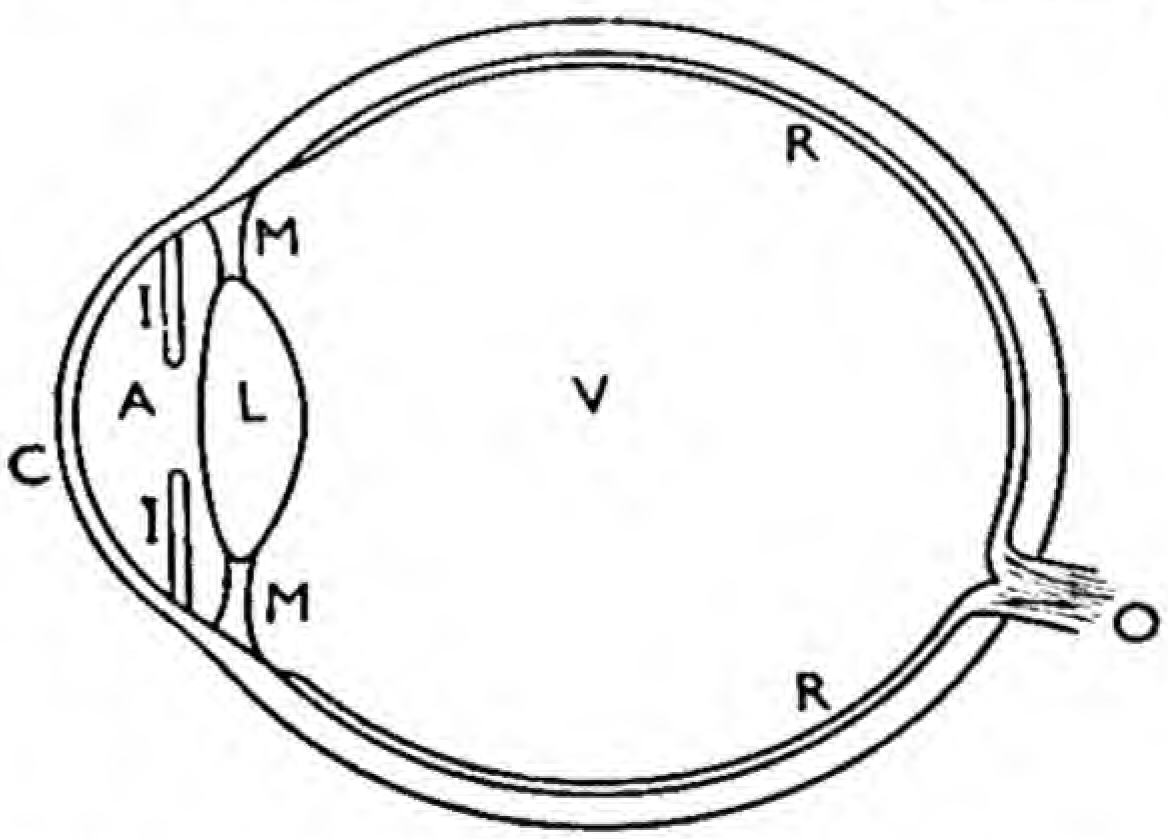


Fig. 49
Diagrammatic Sketch of the Eye
A Aqueous Humour
C Cornea
II Iris
L Crystalline Lens
MM Ciliary Muscle
O Optic Nerve
RR Retina
V Vitreous Humour

C, known as the *cornea*, and a sensitive surface R (the *retina*), on which the lens forms an image of the object to be seen. The *optic nerve*, O, conveys to the brain the effect on the retina caused by the light, and we then see. The *iris*, I, is a stop placed before the crystalline lens which automatically changes the size of its central hole according to the strength of the external light.

The relative positions of the lens and the retina (*i.e.* the distance v) are fixed, so that there should be only one position

of an object (one value of u) for which the image would be formed on the retina, according to equation (3·33). The curvature of the lens surfaces, however, is controlled by a muscle—the *ciliary* muscle, M,—which automatically changes the focal length f when we change the object of our vision to a more or less distant one. This process is known as *accommodation*. The power of accommodation tends to decay with age, and spectacles become necessary. The average focal length of the crystalline lens, however, is small, so that since from equation (3·34) $\frac{1}{v}=\frac{1}{u}+\frac{1}{f}$, the term $\frac{1}{f}$ is dominant in determining the value of v unless u is small. That is to say, for objects beyond a certain distance the image is always formed almost at the principal focus, so that practically no accommodation is needed. This distance is known as the *least distance of distinct vision*. For normal persons it is about 1 foot.

Inversion of Image

It is often thought to be mysterious that we do not see objects upside down, for the image formed on the retina is, of course, inverted (see, for instance, Fig. 47). Vision, however, is determined not only by the image on the retina but also by the mind's interpretation of that image, and it is doubtless in this part of the process that the erection is made. An image is formed on the retina of a dead person, but here presumably no vision at all follows. We have therefore no right to suppose that what we see is represented exactly by the retinal image.

Defects of Vision

The commonest defects of the eyes are "long sight," or *hypermetropia*, and "short sight," or *myopia*. In the former the image of an object at the normal least distance of distinct vision is thrown behind the retina, and in the latter it is thrown in front, *i.e.* in the former the focal length of the crystalline lens is too great, and in the latter it is too small.

The remedy, as may easily be seen from formula (3·39), is to use converging spectacle lenses for long sight and diverging spectacle lenses for short sight. *Astigmatism* also is fairly common. This is a defect arising from uneven curvature of the surfaces of the crystalline lens (*i.e.* the surfaces are not truly spherical), and it is remedied by the use of lenses whose surfaces also have uneven curvature—often that of a cylinder.

The greater the defect of an eye the shorter must be the focal length of the correcting lens. For this reason the *power* of a lens, as it is called, is usually stated by opticians in terms of the reciprocal of the focal length. The unit is 1 *dioptre*, which is the power of a lens having a focal length of 1 metre. Power (unlike focal length) is reckoned positive for converging lenses and negative for diverging lenses.

Photopic and Scotopic Vision

The process of vision after the image has been formed on the retina is the concern of physiology rather than of physics, but a very brief reference to one aspect of it will be useful for our purpose. Vision in bright light is of a different character from vision in faint light ; the former, occurring in the daytime, is known as *photopic* vision, and the latter, occurring at night in the absence of artificial illumination, is known as *scotopic* vision. The difference is associated with the structure of the retina, which is composed of a large number of very small formations known as *rods* and *cones*, from their geometrical shape. The cones operate in photopic, and the rods in scotopic vision. The two processes are distinct, and a person may have good photopic and bad scotopic vision. Thus miners, accustomed to long periods in the dark, sometimes lose the power of photopic vision ; and, on the other hand, the disease of " night blindness," which afflicts people who may be able to see well by day, is well known. It follows, for example, that a person's power of vision by day is an inadequate guide to his potentialities as a night flyer. It is only in photopic vision that we have the power of dis-

tinguishing colours. At night, even in moonlight, colours showing a wide variety by day are distinguished, if at all, only by their relative greyness or blackness.

The Camera

The camera is optically identical with the eye, a sensitized plate or film taking the place of the retina, but since there is no automatic accommodation, objects at different distances are brought into focus by changing the distance from the lens to the plate. There is usually an *iris diaphragm*, which serves the same purpose as the iris of the eye. The chief difference between the eye and the camera is that the image on the retina vanishes without trace when (or soon after) the rays from the object cease to enter the eye, whereas the image on the photographic plate can be preserved permanently if the plate is developed and fixed. Another difference is that the eye cannot see a faint object better (but rather worse) by looking at it a long time, but the photographic plate can store up light and photograph extremely faint objects by long exposure.

The Microscope

The Simple Microscope

One of the most widely used optical instruments is the microscope, which has two forms—the simple microscope, or ordinary magnifying glass, and the compound microscope. The former consists simply of a single converging lens, placed between the eye and the object to be observed and at a distance from the latter less than its focal length. A virtual image is then formed, as in Fig. 50. The eye placed at E sees the object as though it were at I, and the position of the lens is adjusted so that I is at the least distance of distinct vision.

The Compound Microscope

The compound microscope is a much more powerful instrument. It consists of two parts—an *objective* and an *eyepiece*. The objective is a short-focus combination of lenses which

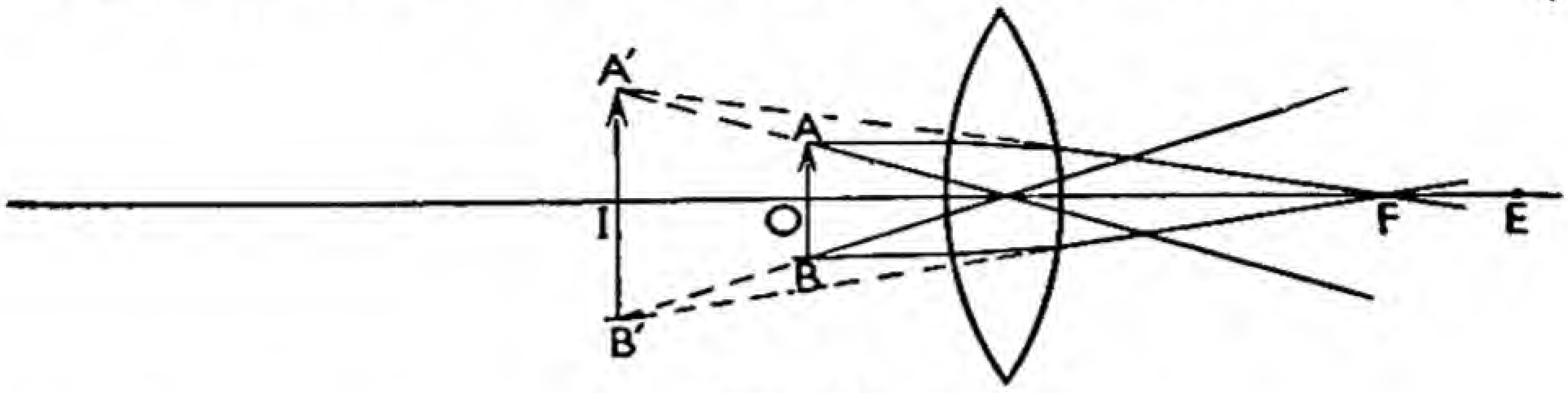


FIG. 50
Formation of Virtual Image A′B′ of Object AB by Simple Microscope

forms a much enlarged real image of the object, and the eyepiece (though usually a combination of lenses rather than a single lens) acts as a simple microscope with this real image as its object, so that the final image is a virtual one. Fig. 51

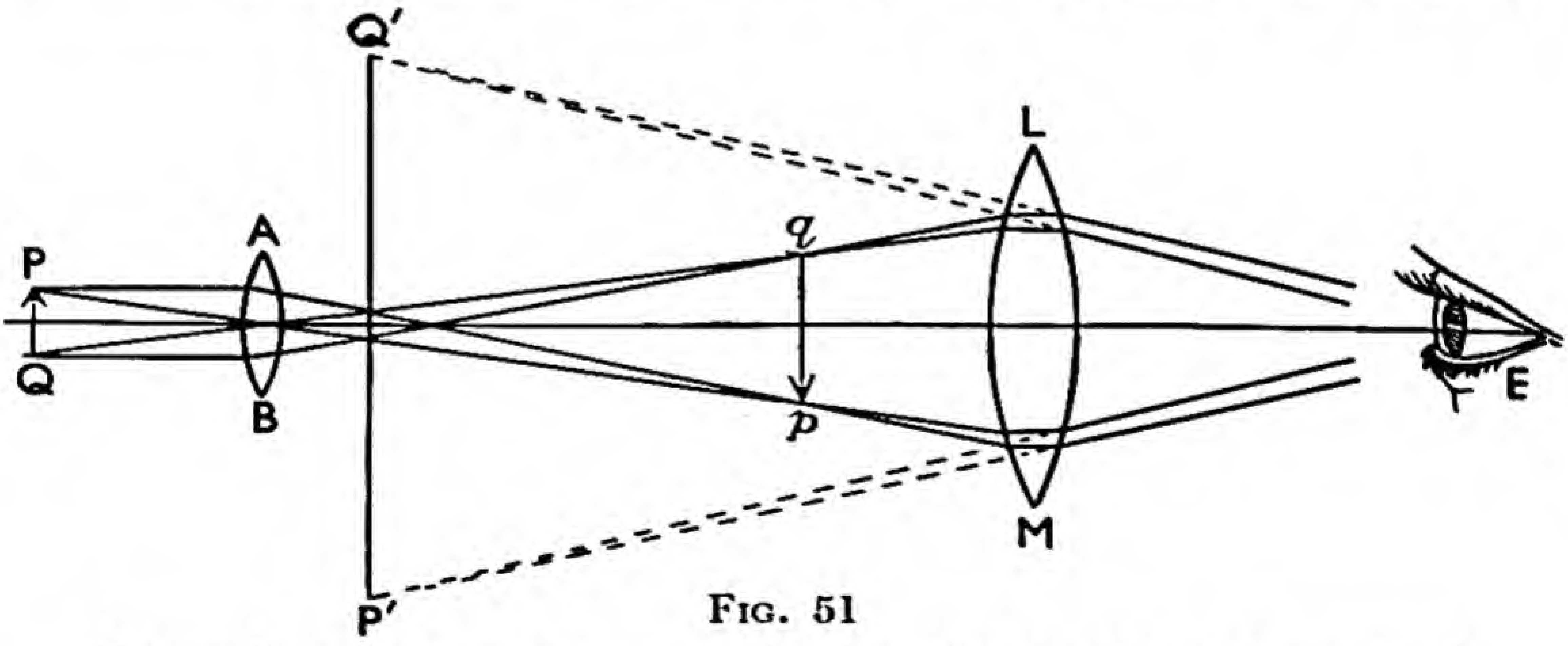


FIG. 51
Formation of Virtual Image Q′P′ of Object PQ by Compound Microscope
AB Objective LM Eyepiece

shows the path of the rays. The object PQ is placed just outside the principal focus of the objective AB, and the real image *pq* is formed. The rays continue onwards to the eye-

piece LM, and thence, after refraction, to the eye E, which sees the enlarged virtual image P′Q′.

The Telescope

The telescope, in all its various forms, consists, like the compound microscope, of two distinct parts—the *object-glass* and the *eyepiece*. The function of the former—a converging lens or lens combination or a concave mirror—is to collect light from the object and bring it to a focus as a real image. The eyepiece, like the microscope eyepiece, magnifies this image. If the telescope is to be used to photograph an object, a photographic plate receives the image formed by the object-glass, and no eyepiece is used.

The Astronomical Telescope

In the "astronomical" telescope (used for some terrestrial purposes also) the object is practically at infinity, and the

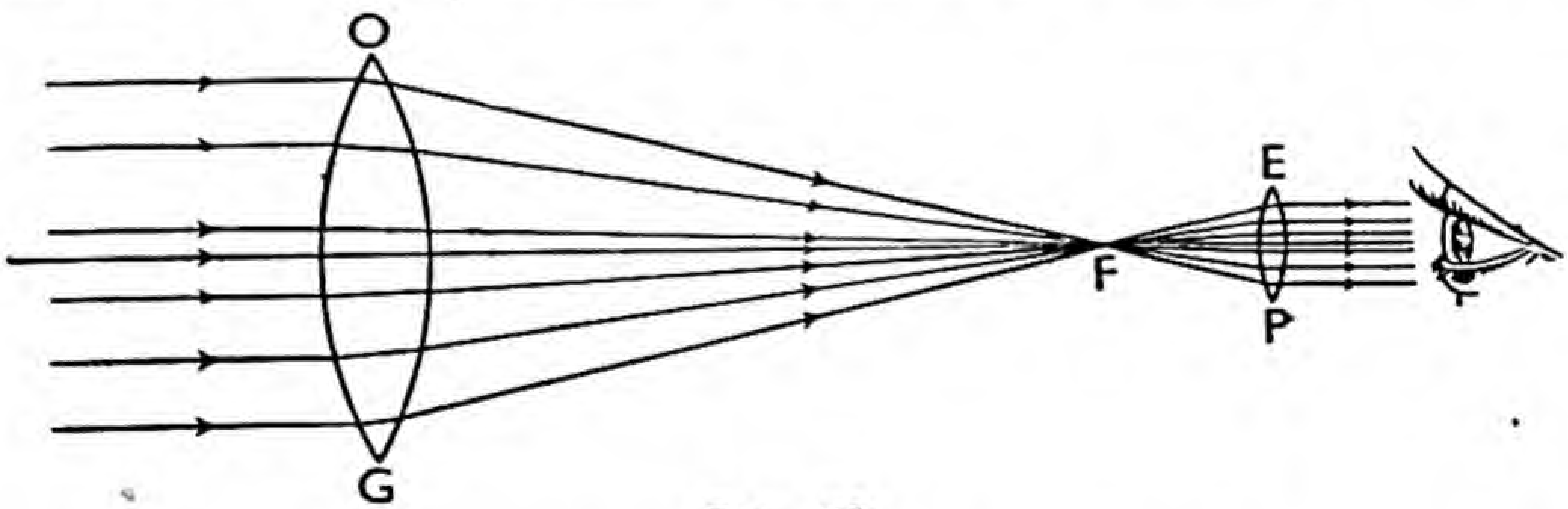


Fig. 52

Path of Rays from Distant Point through Astronomical Telescope
OG Object-glass EP Eyepiece F Principal Focus of OG and EP

first image is formed at the principal focus F of the object-glass (Fig. 52). The eyepiece (of much smaller focal length) is then placed so that F is at its focus also. The rays from each point of the object then issue parallel, and the eye sees an inverted image "at infinity." The magnifying power is defined as the ratio of the angle subtended at the eye by the

image to that subtended by the object. Fig. 52 is drawn for a point object, which subtends no angle at the eye (this is effectively so, for instance, when a star is observed), but for an object of finite angular size it can be shown that the magnification is given by the ratio of the focal length of the object-glass to that of the eyepiece.

Parallax: To facilitate measurements, *crosswires*—consisting of two spider threads bisecting one another at right angles—are placed at F. They are fixed in a draw tube which can be moved in or out until the image formed by the object-glass coincides with them. To ensure that this is so, the eyepiece is focused on the crosswires by a separate movement, and then eyepiece and crosswires together are moved in and out until there appears to be no relative movement between the image of the crosswires and that of the object when the eye is moved from side to side. If the two images are not at the same position in space they will appear to separate when the eye makes this movement, but if they are at the same position they will keep together for all positions of the eye. There is then said to be no *parallax* between them.

Reflectors and Refractors: Telescopes used in astronomy often have an object-glass consisting of a concave mirror instead of a lens. They are then known as *reflectors*. Their principle is precisely the same as that of the *refractors* already described.

The Galilean Telescope

In many terrestrial telescopes, and in field and opera glasses, a *diverging* lens is used for the eyepiece, giving the *Galilean* telescope, as it is called. This has the advantages, first, that it gives an erect instead of an inverted image, and secondly, that it enables the telescope to be shortened, for the eyepiece is placed between the object-glass and its image. Thus, in Fig. 53, OG by itself would form an image at F, but the eyepiece EP, placed so that F is at its principal focus, makes the rays emerge parallel, and the eye again sees a virtual

image at infinity. Such a telescope, of course, cannot be used to photograph the image of an object. The magnifica-

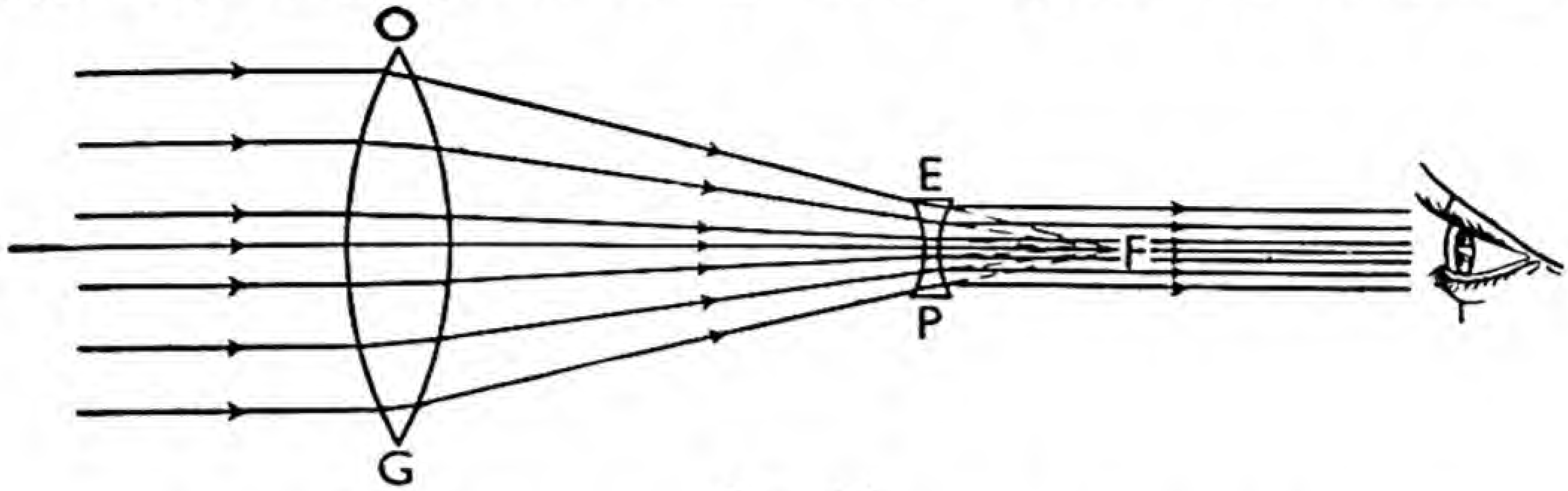


FIG. 53
Path of Rays from Distant Point through Galilean Telescope
OG Object-glass EP Eyepiece F Principal Focus of OG and EP

tion is again given by the ratio of the focal lengths of object-glass and eyepiece.

THE GONIOMETER

The goniometer is an instrument for measuring angles. The angle between two reflecting surfaces is most accurately measured by optical methods. We shall take as an illustration the angle between two faces of a triangular prism. A beam of light, made parallel by a lens, is allowed to fall on the two faces as shown in Fig. 54, and is reflected from them into a telescope which can be brought to view each reflected beam in turn. Let α be the angle to be measured, and β the angle between the positions of the telescope when it receives the two reflected beams. It is easy to show that $\beta = 2\alpha$.

For, in Fig. 54, we see that AML is equal to BMN, from the law of reflection, and is also equal to MAR, since MA cuts the parallel lines LM and TR. Hence BMN = MAR. Similarly, CHG = HAR. Hence

$$\text{BMN} + \text{CHG}(= \beta - \alpha) = \text{MAR} + \text{HAR}(= \alpha)$$

$$i.e.\ \beta = 2\alpha.$$

The method of measuring β consists in placing the prism on a table and illuminating it in the manner shown, by light from a point or line source. A telescope, which can be rotated about a vertical axis through A and whose position

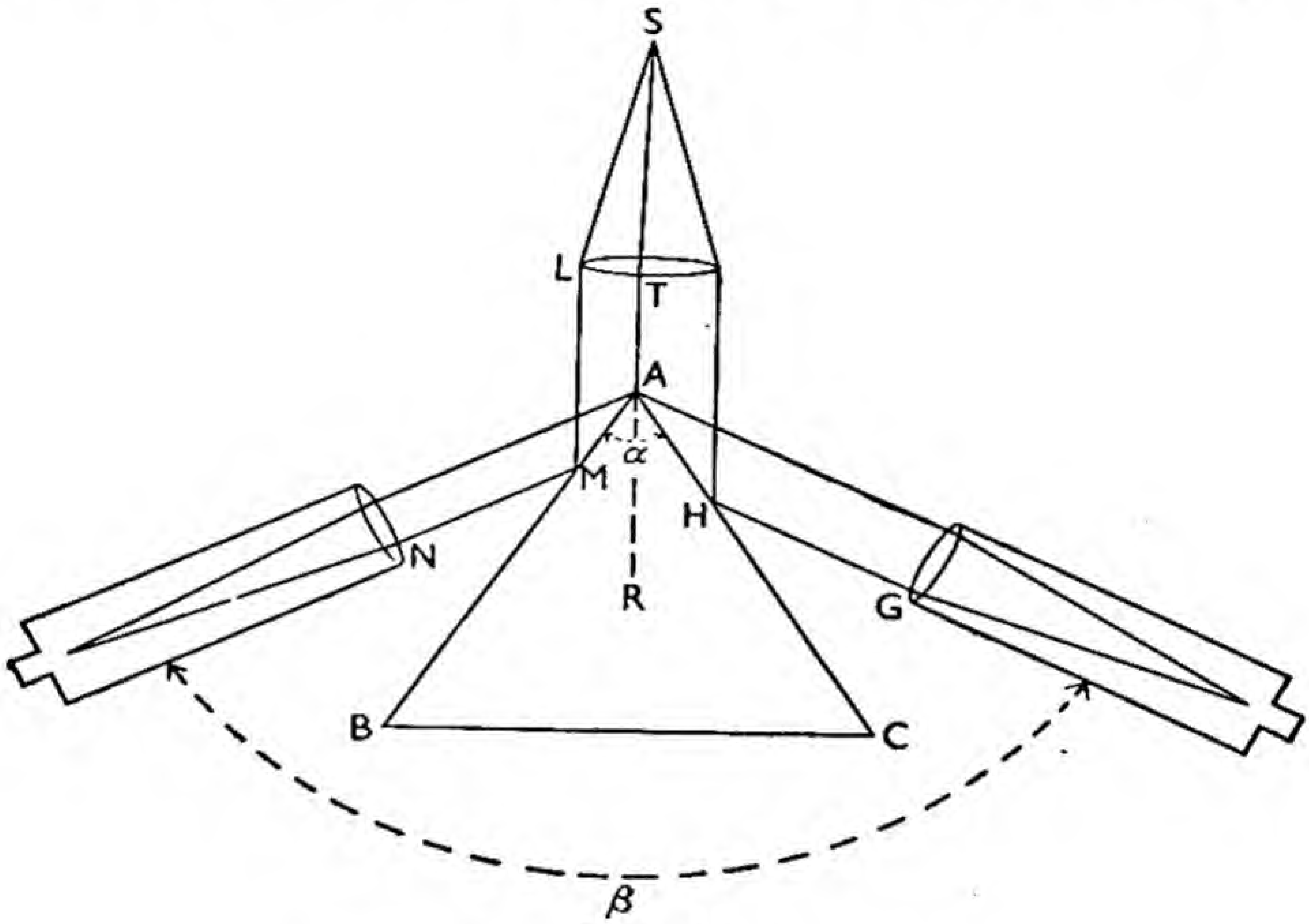


FIG. 54

Optical Measurement of Angle α

is indicated by a scale of degrees, is focused on the images reflected from the two faces, and the angle between the two positions of the telescope is the angle β. The angle of the prism is half this. The instrument used is generally called a *spectrometer*, because of its use also for measuring the positions of spectrum lines produced by refraction through the prism.

THE VELOCITY OF LIGHT

Whatever light may be, it can be shown that it does not travel instantaneously but takes a definite time to move from one point to another. One method of showing this, and

of measuring the velocity, will suffice. In Fig. 55, W is a wheel having at its edge regularly arranged teeth, behind one of which a point source of light is placed. The wheel can be rotated, so that the light can either shine through a space between neighbouring teeth or be obstructed by a tooth, and, as rotation goes on, these conditions alternate. When the light passes through, it falls normally on a plane mirror M from which it is reflected back along its path. But during the double journey the wheel has rotated, so that, on returning, the light may encounter a tooth and be absorbed,

Fig. 55
Apparatus for Measuring the Velocity of Light
S Source of Light W Rotating Toothed Wheel M Plane Mirror

or it may pass through a space. If, now, we look from behind the wheel towards the mirror and gradually increase the speed of rotation, we shall at first see light intermittently, for, since its speed is exceedingly great, the light which goes through a space will travel to M and back before the wheel has moved an appreciable amount, and so will get through the same space again to our eye. A speed of rotation will come, however, when, during the passage of the light from W to M and back, a tooth manages to reach the place where the space was which let the light through. The returning light will then be absorbed, and we shall not see it. If this happens for one space it will happen for all if the rotation

is uniform, and the eye will therefore cease to see the light altogether.

Measurement of Velocity of Light

Suppose the speed of the wheel when this happens is n revolutions (*i.e.* an angle of $2\pi n$ radians) per second, and suppose there are m teeth (*i.e.* the number of teeth and spaces together is $2m$). Then the angle moved through by a tooth in getting to the adjacent space is $\frac{2\pi}{2m}$, and the time taken by the movement is $\frac{2\pi}{2m} \times \frac{1}{2\pi n} = \frac{1}{2mn}$ seconds. Now during this time the light has travelled from W to M and back—a distance $2l$ cm., say. The velocity of the light is therefore $2l$ cm. in $\frac{1}{2mn}$ seconds, *i.e.* $4lmn$ cm./sec. By taking a long distance for l, it is not difficult to make a wheel with sufficient teeth and to rotate it rapidly enough to measure the velocity of light, in spite of the fact that light travels about 3×10^{10} cm. (186,000 miles) a second in empty space, and very slightly more slowly in air.

Refractive Index and Velocity of Light

The velocity of light is of little direct importance in terrestrial observations, but it is of fundamental significance in astronomy and in physical theory. It is indirectly important, however, because, as already implied, the velocity is different in empty space and in a transparent material body, and differs also in different material bodies. The refractive index, in fact, is the ratio of the velocities of light in the two media concerned. Thus, if light passes from medium 1 to medium 2, then

$$_1\mu_2 = \frac{c_1}{c_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4 \cdot 1)$$

where c_1 and c_2 are the respective velocities; and when

the first medium is a vacuum (or, in practice, air), we have

$$\mu=\frac{c}{c_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4\cdot2)$$

where c is the velocity of light *in vacuo*. c_2 is always less than c, *i.e.* all material media are optically denser than a vacuum.

EXERCISES

1. What is meant by the " least distance of distinct vision " ? Explain why vision may be distinct at greater distances but not at smaller. If this distance for a certain person is 2 feet, and he wishes to read comfortably at a distance of 1 foot from the eye, find the type and focal length of the spectacle lenses which he must use.

2. Describe the astronomical telescope. If the focal lengths of the object glass and eyepiece are respectively 10 feet and 6 inches, what is the angular aperture of the image seen in the telescope of an object 6 feet high at a distance of $\frac{1}{4}$ mile from the observer ?

3. Describe a method of measuring the angle between two plane glass surfaces.

4. What evidence is there that light takes a finite time to travel from one point to another ? How may its velocity be measured ?

CHAPTER V

PHOTOMETRY

Measurement of Intensity of Light

THE problem of measuring the intensity of light is difficult, both theoretically and practically. The theoretical difficulties arise from the fact that the conception of light is rather indefinite : the limits of the visible spectrum (see p. 28) vary for different people, so that what is light for one may be darkness for another. Further, there are two distinct principles on which we can measure the brightness of a luminous body. We can measure it in terms of the amount of light sent out per second—a *dynamic* measure—or in terms of the brightness of its appearance as compared with an arbitrary standard of luminosity—a *static* measure. Again, there is a related problem in the measurement of the intensity of illumination, by some external source, of a non-luminous body. This is clearly a different matter from the measurement of the brightness of a source of light.

Illuminating Power

Consider first a point source of light. Its brightness is measured by its *illuminating power*, which is defined as the light emitted by the source in unit time, divided by 4π. The light emitted in unit time should theoretically be measured in absolute units, as the number of ergs of light energy so emitted, but this is very difficult, since light is always accompanied by radiations outside the visible spectrum, which also are forms of energy, and any simple method of measuring the energy of light (such as absorbing it by a lampblack surface—

see I, 185—and measuring the rise of temperature produced) makes no distinction between the visible and invisible radiations. For this reason an arbitrary standard of luminosity is chosen for practical purposes. It is known as the *standard candle*, and the source is a sperm candle, six of which weigh a pound, burning at the rate of 120 grains of wax an hour. The intensity of the light given is called 1 *lumen*. The actual standards generally used are electric lamps which have been compared with the standard candle, so that their emission in lumens is known. The illuminating power of a point source of light is then the number of lumens which it emits divided by 4π.

If we have a surface instead of a point emitting light, we measure its luminosity by the *surface brightness*, which is the light emitted per unit area per second.

Intensity of Illumination

The degree to which a non-luminous body is illuminated by an external source is measured by the *intensity of illumination*, which is the amount of light falling normally on unit area in unit time. This, of course, is not a measure of how bright the body appears, because, although the same amount of light per unit time may fall on a black and on a white body, the latter will appear much brighter than the former since it scatters the light received into space again.

The Inverse Square Law

It can easily be proved that unit intensity of illumination is the illumination of a surface at unit distance from a source of unit illuminating power. For, consider a source, of illuminating power L (Fig. 56), at the centre of a sphere of radius r. All the light emitted per second (*i.e.* $4\pi L$) falls normally on the inner surface of the sphere, whose area is $4\pi r^2$. Hence the intensity of illumination of the surface is

$$4\pi L \div 4\pi r^2 = \frac{L}{r^2} \quad . \quad . \quad . \quad . \quad . \quad (5\cdot1)$$

so that if $L = 1$ and $r = 1$, the intensity of illumination is unity.

The expression (5·1) is very important. It shows that the intensity of illumination of a surface falls off as the square of its distance from a luminous point. One result of this is that the *apparent* surface brightness of an illuminated area is the same from whatever distance it is seen. For, suppose the distance is doubled. Then the amount of light received from the surface (regarding it as a source of the light which it scatters towards the eye) is only one quarter of that received before. But the area also is apparently reduced to one quarter, so that the light received from each unit of angular area is unchanged. If the distance is so great that the area appears to be merely a point, this ceases to be true, for we are then dealing effectively with a point source, and surface brightness ceases to have a meaning. An astronomical illustration of this is that the surface brightness of a nebula (luminosity per square degree) is independent of the distance of the nebula, but the brightness of a star varies inversely as the square of its distance.

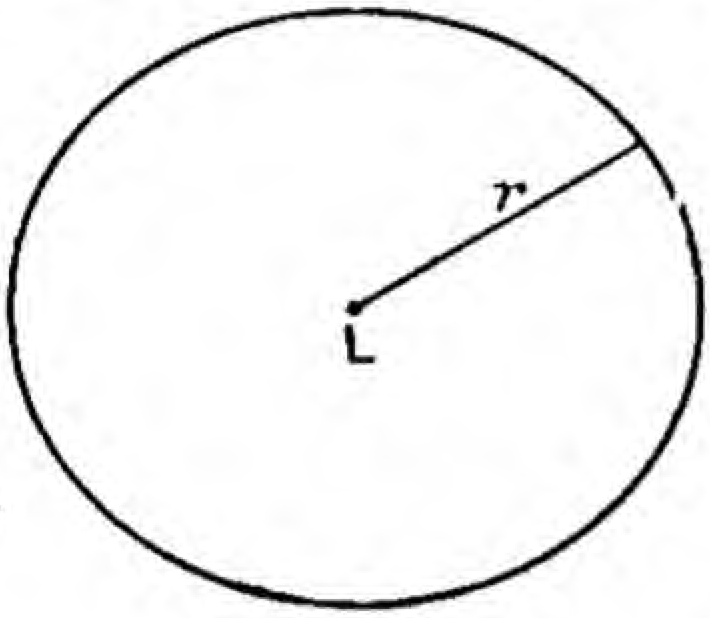


FIG. 56
Diagram illustrating the Intensity of Illumination at a Distance r from a Source of Illuminating Power L

The unit of intensity of illumination is the intensity of illumination at unit distance (1 cm.) from a standard candle, *i.e.* from a source of 1 lumen. This intensity is called 1 ***phot***. The intensity at 1 *metre* from a standard candle is often used as an alternative standard. This is clearly 10^{-4} phot, and is called 1 *lux*. Other units also are used for special purposes.

Photometers

An instrument for measuring luminosity is called a ***photometer***. There are many different kinds of photometer, and in most of them the source to be measured is compared with a standard

lamp and the luminosity is calculated by the use of equation (5·1). A general difficulty is that the source is often of a different colour from the standard, and it is not easy to judge when sources of different colour are equal in intensity. An attempt to overcome this difficulty is sometimes made by the so-called "cascade" method. A series of equal standard lamps is prepared, in which the colour-change from one to the next is very slight, so that accurate comparison can be made. The whole series runs right through the colours of the spectrum. The intensity of the source to be measured is

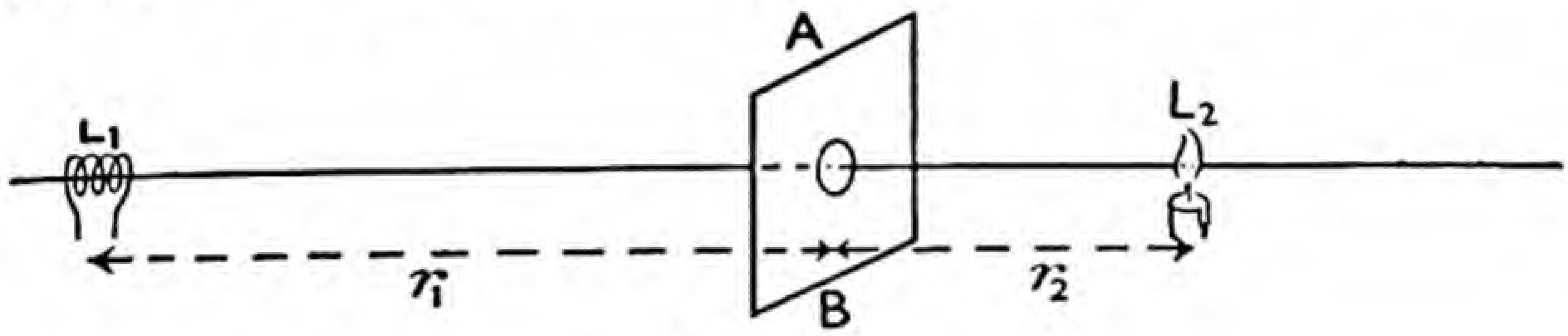


FIG. 57

Bunsen's Grease-spot Photometer

AB White Card with central grease spot
L_1 Source to be measured, distant r_1 from AB
L_2 Standard Source, distant r_2 from AB

then compared with that of the standard nearest to it in colour.

Bunsen's Grease-spot Photometer : As an example of the general principle of photometry, Bunsen's "grease-spot" photometer will serve very well. A white card with a central grease-spot is placed between the two sources, of illuminating power L_1 and L_2, to be compared (Fig. 57). When the sources are at distances r_1 and r_2, respectively, from the card, the intensities of illumination on the two sides are $\frac{L_1}{r_1^2}$ and $\frac{L_2}{r_2^2}$ respectively. Now the light falling on the card is partly transmitted and partly reflected (the amount absorbed by a white card may be neglected). Suppose the grease-spot transmits a fraction α, and the rest of the card a fraction β, of the incident light ; then the corresponding fractions re-

flected must be $(1-\alpha)$ and $(1-\beta)$. An observer of the right-hand side of the card will then receive from the grease-spot an amount of light proportional to $\frac{L_2}{r_2^2}(1-\alpha)+\frac{L_1}{r_1^2}\alpha$, and from the rest of the card, an amount proportional to $\frac{L_2}{r_2^2}(1-\beta)+\frac{L_1}{r_1^2}\beta$. These quantities can be varied by varying the distances r_1 and r_2. Let these distances, then, be adjusted until the grease-spot and the rest of the card appear equally bright, so that the boundary circle between them disappears. We shall then have

$$\frac{L_2}{r_2^2}(1-\alpha)+\frac{L_1}{r_1^2}\alpha=\frac{L_2}{r_2^2}(1-\beta)+\frac{L_1}{r_1^2}\beta \quad . \quad . \quad . \quad (5\cdot2)$$

$$\left.\begin{aligned} \text{from which} \quad & \frac{L_1}{r_1^2}=\frac{L_2}{r_2^2} \\ \text{or} \quad & \frac{L_1}{L_2}=\frac{r_1^2}{r_2^2} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5\cdot3)$$

The experiment therefore consists in moving the card to such a position between the fixed sources that the grease-spot is no longer distinguishable from the rest. The relative intensities of the sources are then given by equation (5·3), and if L_2 is a standard lamp of known luminosity, the illuminating power of L_1 is known.

The Lummer-Brodhun Photometer: A more accurate instrument is the Lummer-Brodhun photometer, illustrated in Fig. 58. L_1 and L_2 are placed on opposite sides of an opaque white screen AB at distances great compared with the size of the screen, so that the light falling on the screen from each side may be regarded as effectively parallel. This light is scattered in all directions, and some falls on the mirrors M_1 and M_2, from which it is reflected to the pair of prisms PQ, the edges of one of which are slightly bevelled, as shown in the figure.

Consider the light falling from M_2 on this system. That

falling opposite the bevelled edges of P is totally reflected, its angle of incidence being greater than the critical angle, so that an eye placed at E will receive it ; but the light in the centre meets no glass/air surface until it reaches the farther side of P, where it falls normally and goes through with only a very small loss by reflection, none of it, therefore, entering

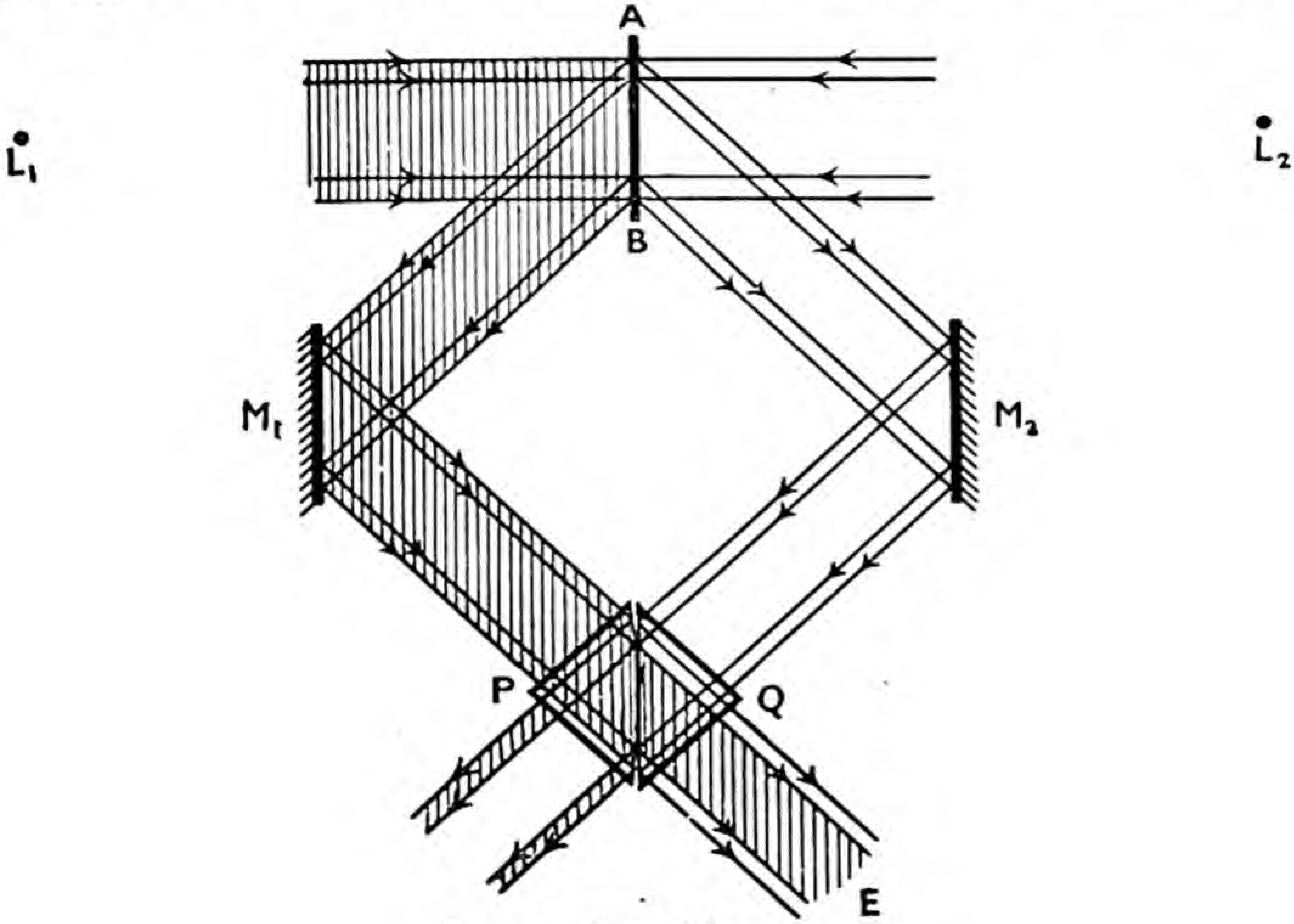


FIG. 58
The Lummer-Brodhun Photometer

the eye. The light from M_1 behaves similarly, the central portion only travelling straight through towards E. What the eye sees, therefore, is, in the centre of the field of view, the light from L_1 via M_1, and surrounding this, a band of light from L_2 via M_2. By adjusting the distances of L_1 and L_2 from AB, these two areas of light can be made equal in intensity, and when this happens the two sides of AB are equally illuminated. The ratio $\frac{L_1}{L_2}$ is then equal to $\frac{r_1^2}{r_2^2}$, where

r_1 and r_2 are respectively the distances of L_1 and L_2 from AB.

Photo-Electric Photometers: In much modern work *photo-electric* photometers are used (see p. 250). The intensity of a light source is measured in these instruments by the number of electrons it is able to release from a metal surface. The details are beyond our scope.

EXERCISES

1. Discuss the difficulties of defining precisely and usefully what is meant by the intensity of the light radiated by a given body, and state how they have been overcome.

2. Describe an accurate form of photometer, and explain how, by subsidiary experiments if necessary, it can be used to compare the brightness of two sources of light of different colour.

3. Two sources of light illuminate a screen equally when one is three times as far from it as the other. If the nearer, originally at a distance of 1 foot, is removed 6 inches farther from the screen, by what factor does the illumination produced by the other exceed that which the former now gives, and how far must the brighter source be withdrawn to restore equality of illumination?

4. Two street lamps of equal brightness, each 12 feet above the ground, are situated on a road 300 feet in length at points distant 100 feet and 200 feet respectively from one end. Draw a curve showing the variation of illumination on the ground along the length of the road.

CHAPTER VI

DISPERSION

Colour and Refraction

So far it has scarcely been necessary to mention the fact that light has various colours, for almost all that has been said applies to light of any colour. We must now, however, consider phenomena whose characteristics depend on the colour of the light concerned.

We may say at once that the law of reflection is unaffected; light of all colours is reflected along the same path if it is incident along the same path. The law of refraction, however, is modified, for although Snell's law is true for each colour separately (*i.e.* for light of any one colour, the fraction $\frac{\sin i}{\sin r}$ is constant for all values of i) the value of $\frac{\sin i}{\sin r}$—the refractive index—varies with the colour.

Colour and Wave-length

In dealing with these questions the wave theory of light is most generally useful, and on this theory different colours are distinguished by their different wave-lengths or frequencies. We shall therefore generally speak of wave-length rather than colour, as being a more definitely measurable entity. We have already said (p. 28) that visible light is regarded as consisting of transverse waves ranging from about 8×10^{-5} to 4×10^{-5} cm. The unit generally used in measuring these wave-lengths is called the *angstrom*: it is equal to 10^{-8} cm. Visible light, then, ranges from 8,000 to 4,000 angstroms (written 8,000A. — 4,000A.). The longer waves

are at the red end of the spectrum and the shorter waves at the violet end.

Now the refractive index of a medium steadily increases as the wave-length diminishes (except in certain " anomalous " media, which we shall ignore) ; the shorter waves are on this account said to be *more refrangible* than the longer ones. This means, of course, that the diminution of velocity when light enters a denser medium is less for the longer waves than for the shorter ones.

Analysis of Light

It is a well-known fact that when lights of different colours are mixed together we see a single resultant colour in which we cannot distinguish the individual components. In this respect light is unlike sound, for we can distinguish the separate notes in a chord of music. It follows that when we see light of a particular colour, we have no means of knowing without experiment whether it consists of light of a single wave-length or of a mixture of wave-lengths. In order to find out we must analyse the light, and the fact that the different colours are refracted differently at once suggests a method ; we may allow the light to fall obliquely on a refracting surface, so that the component colours will emerge in different directions.

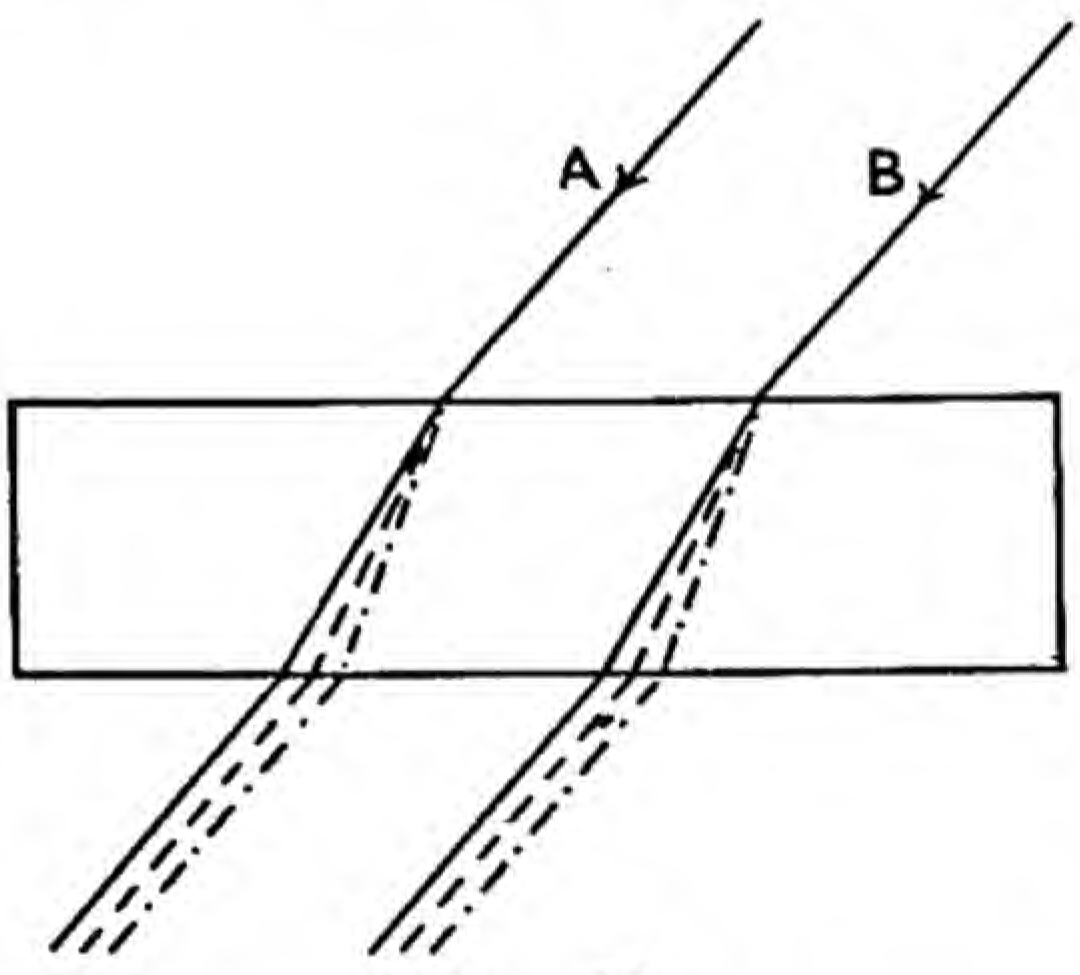


FIG. 59
Refraction of Rays of Different Colours through Parallel-sided Block of Glass

In carrying out this experiment we cannot use a parallel-sided block of glass, for although the rays pursue different paths in the glass, they will all take up their original direction on emerging into the air again. This is illustrated in Fig. 59,

which shows a beam of light (of which only the boundary rays, A and B, are drawn) consisting of three colours passing through such a block. It will be seen that there is only a slight lateral separation, the three colours travelling in the same direction when they come out into the air again.

We may, however, obtain a permanent separation by using a prism. Fig. 60 shows a similar beam passing through a prism, and here it is clear that the rays emerge into the air

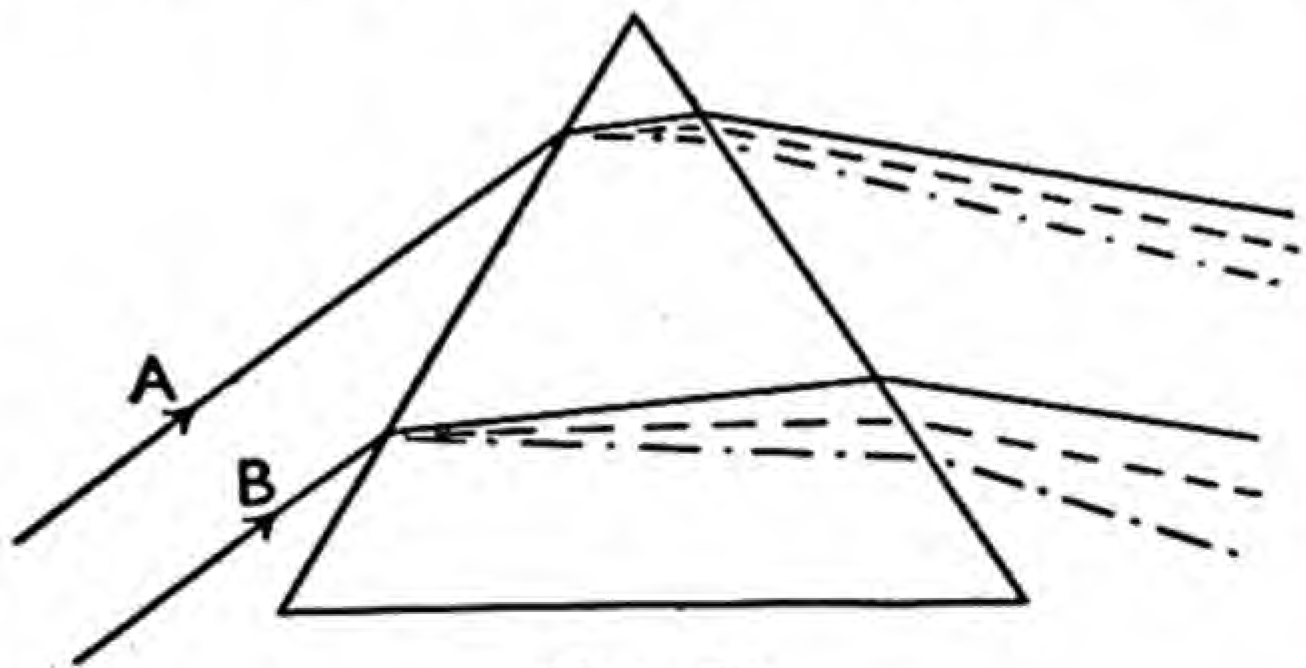


FIG. 60
Dispersion of Rays of Different Colours by Triangular Prism

in different directions. Each colour, in fact, forms its own virtual image of the source in the manner described on pp. 72–73.

The Spectroscope

This is the principle of the *spectroscope*, a very important optical instrument. To get the best results we use a narrow source (to prevent overlapping of the images), obtained by passing the light through a fine slit. The light is then made parallel (by placing the slit at the principal focus of a converging lens) so that it all falls on the prism at the same angle, thus ensuring that the images formed shall be free from aberration. Light of each colour then remains parallel during its journey through the prism and beyond, where it is received by a telescope of astronomical type, which brings it

to a focus as a set of *real* images of the illuminated slit, one for each colour in the incident light. This is shown in Fig. 61. The tube AB, containing the slit and lens, is called the *collimator*. It is customary to set the prism at the angle of minimum deviation for one of the colours, so that if it is

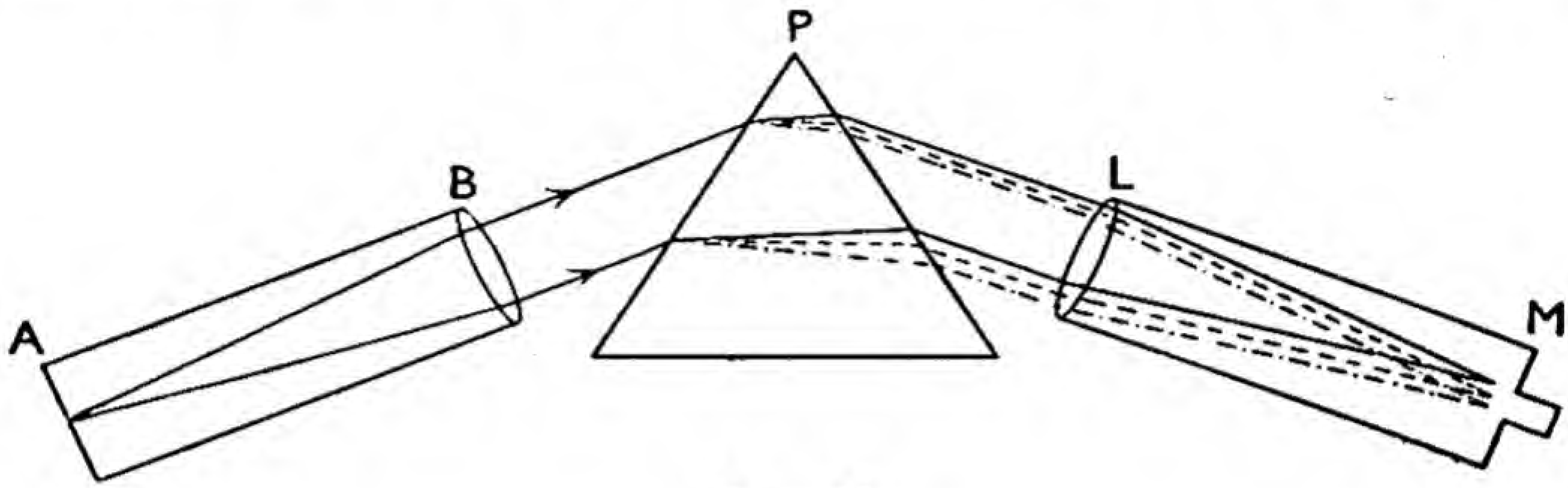


FIG. 61
The Spectroscope
AB Collimator P Prism LM Telescope

accidentally disturbed its position can be recovered again. This position, however, has in general no other advantage.

If it is desired to photograph the spectrum, the eyepiece of the telescope is removed, and the images formed by the object-glass are received on a photographic plate; the telescope, in fact, is converted into a camera. The instrument is then known as a *spectrograph*.

Spectrum Analysis

Ordinary " white " light, such as that from an electric glow lamp, when analysed by the spectroscope, is found to contain all the colours of the visible spectrum. Instead of three isolated colours, as in Fig. 60, we have a continuous band of colour varying gradually from red at one end, through orange, yellow, green, blue, indigo, to violet at the other. This is known as a *continuous spectrum*, and it is obtainable from any brightly glowing solid or liquid. Glowing gases, however, give a number of isolated colours, each appearing as an

image of the slit ; they are called *spectrum lines* on that account. No two gases, of different chemical composition, give the same spectrum. This is the basis of *spectrum analysis*, which consists in vaporizing the substance to be analysed, making it luminous by heat or by an electric current or otherwise, and examining its radiation by the spectroscope. The resulting spectrum is always characteristic of the materials present, and if there are several of them their spectra, under suitable conditions, all appear in the resultant analysed light. Fig. 62 is a photograph of the spectra of iron, calcium, and strontium. It will be seen that they differ from one another, the positions of the lines being dependent on the wave-lengths.

The Solar Spectrum

The spectrum of sunlight is neither a pure continuous nor a pure line spectrum. It consists of a continuous spectrum crossed by a number of relatively *dark* lines (called *Fraunhofer* lines). This is due to the fact that the surface of the Sun radiates light which would give a continuous spectrum, but this light, before reaching the Earth, has to pass through the Sun's and the Earth's atmospheres, which absorb some of the wave-lengths, and dark lines therefore appear where these wave-lengths would otherwise be. It appears that a gas can absorb the same wave-lengths that it can emit, so the wave-lengths of the dark lines can be used to analyse the Sun's and Earth's atmospheres, just as though they were bright lines. In this way we find that the composition of the Sun is very similar to that of the Earth.

Atoms and Spectrum Emission

The process of emission of light has already been described (pp. 22–23). We have seen that an atomic electron can revolve in a large number of different orbits, and when it falls from an outer to an inner orbit it gives out energy in the form of a light-wave. Now a large number of inward transitions are possible, for the electron can fall back from

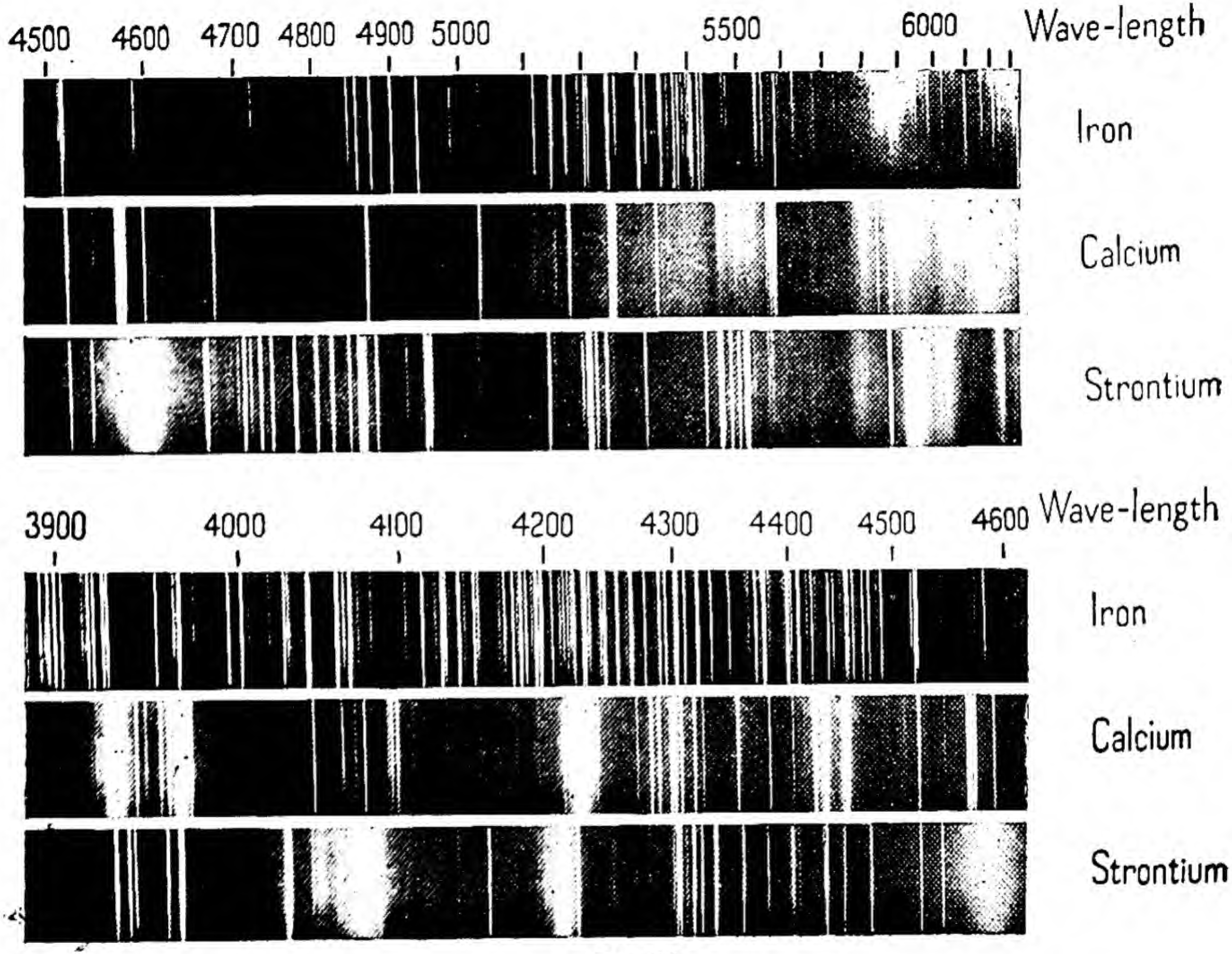


FIG. 62

Spectra of Iron, Calcium and Strontium. (The wave-lengths are in angstroms)

any of its orbits to another (with certain exceptions which we may ignore here). It is found that each such transition results in the emission of a particular wave-length of light, and the whole spectrum consists of all the wave-lengths so emitted. In solids and liquids the atoms are practically in contact, and interfere with one another, so that the energy differences between the orbits vary, and a general confusion of light of all wave-lengths is emitted ; *i.e.* we have a continuous spectrum. In a gas, however, where the activities of each atom are little disturbed by other atoms, the wave-lengths are emitted separately, and appear as lines as in Fig. 62.

Absorption of light takes place when light falls on an atom and removes an electron from an inner to an outer orbit. The light then disappears, and its energy is converted into potential energy of the electron. The energy so transformed is, of course, equal to the difference in the equilibrium amounts of energy of the electron in the two orbits concerned. If the electron falls back to the same orbit again, the energy is reconverted into light of the same frequency or wave-length. In fact, equation (1·1) can be applied to either absorption or emission of light, and that is why the wave-lengths which an atom can absorb are the same as those which it can emit.

Dispersive Power

The separation of the component wave-lengths of a beam of light is known as *dispersion*. It leads us to reconsider all that we have said about refraction, for as the light we use is generally sunlight or the light of electric glow lamps, and this consists of a large number of different wave-lengths, we must regard refractive index μ, as a variable instead of a constant quantity. Different materials vary in their power of separating the colours. The *dispersive power* of a material is accurately measured by the quantity $\frac{d\mu}{d\lambda}$, and it varies with the wave-

length λ. For practical purposes, however, we often take an average value, defined by

$$D = \frac{\mu_v - \mu_r}{\mu_y - 1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6\cdot1)$$

where μ_v, μ_r, and μ_y are the refractive indices for particular wave-lengths of violet, red, and yellow light respectively.

Chromatic Aberration

In dealing with lenses (p. 83) we saw that the focal length f was given by

$$\frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad . \quad . \quad . \quad . \quad (6\cdot2)$$

Since μ is different for the different colours, we see that parallel light is brought not to a single focus but to a series

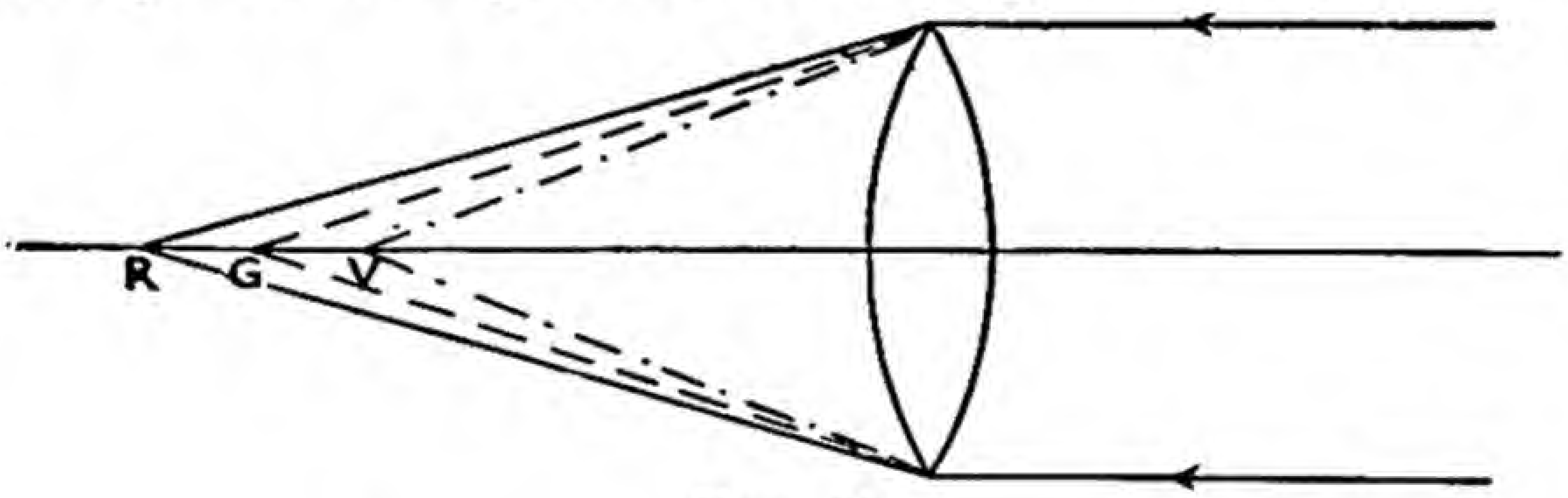


FIG. 63

Chromatic Aberration of a Lens

R Principal Focus for Red Rays G Principal Focus for Green Rays
V Principal Focus for Violet Rays

of foci, and the image of an axial point source of white light is not a point but a line of coloured light lying along the axis (Fig. 63). This phenomenon is known as *chromatic aberration.* The amount of chromatic aberration shown by a lens can be measured as follows. Our formula, applied to the red and violet rays separately, gives

$$\left.\begin{aligned} \frac{1}{f_r} &= (\mu_r - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \\ \text{and} \quad \frac{1}{f_v} &= (\mu_v - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (6\cdot3)$$

M. Sc. Final

$$\text{hence}\quad \frac{f_v}{f_r} = \frac{\mu_r - 1}{\mu_v - 1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6{\cdot}4)$$

and since $\mu_v > \mu_r$ it is clear that f is numerically greater for the longer (red) waves. The quantity $f_r - f_v$, which is easily seen to be given by

$$f_r - f_v = \frac{\mu_v - \mu_r}{\mu_r - 1} \cdot f_v \quad . \quad . \quad . \quad . \quad . \quad (6{\cdot}5)$$

is the measure of the chromatic aberration of the lens.

Fortunately, chromatic aberration can be removed by combining together lenses of different kinds of glass having different dispersive powers. Let us apply formula (3·39) to two such lenses, of focal lengths f and f' respectively, and let F be the focal length of the combination. For red light we have then

$$\frac{1}{F_r} = \frac{1}{f_r} + \frac{1}{f'_r} \quad . \quad . \quad . \quad . \quad . \quad (6{\cdot}6)$$

and for violet light,

$$\frac{1}{F_v} = \frac{1}{f_v} + \frac{1}{f'_v} \quad . \quad . \quad . \quad . \quad . \quad (6{\cdot}7)$$

Hence, in order that F_r and F_v shall be equal (*i.e.* the combination shall have the same focal length for red and violet light) we must have

$$\frac{1}{f_r} + \frac{1}{f'_r} = \frac{1}{f_v} + \frac{1}{f'_v} \quad . \quad . \quad . \quad . \quad . \quad (6{\cdot}8)$$

$$\text{i.e.}\quad \frac{1}{f_v} - \frac{1}{f_r} = \frac{1}{f'_r} - \frac{1}{f'_v} \quad . \quad . \quad . \quad . \quad . \quad (6{\cdot}9)$$

It is possible to choose lenses satisfying this condition because converging lenses have negative and diverging lenses positive focal lengths. Hence, although f_r is always *numerically* greater than f_v, for converging lenses it is *algebraically* smaller, so that we must combine a converging and a diverging lens. They are usually made of crown and flint glass respectively.

A combination satisfying condition (6·9) is called an *achromatic* combination. A common form is illustrated in Fig. 64. Whenever a sharply defined image is required in

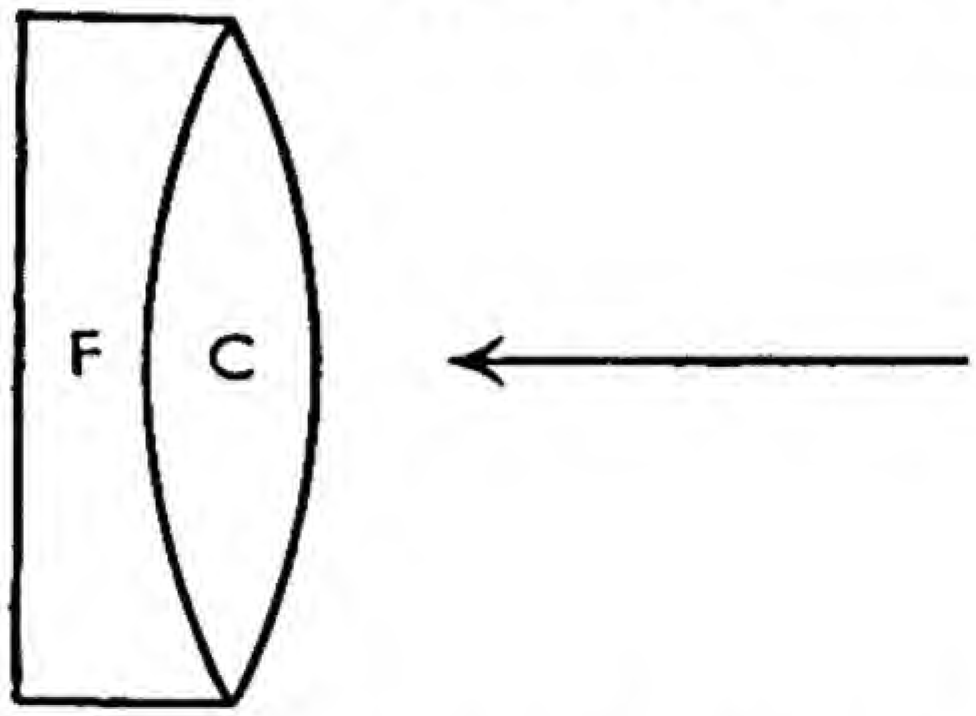


FIG. 64
Achromatic Lens Combination
C Converging lens of crown glass
F Diverging lens of flint glass
Arrow shows direction of light when lens is used as telescope object-glass

light of different colours, such a combination must be used. No combination is perfectly achromatic, for although the red and violet rays may come to the same focus, the intermediate colours do not necessarily do so. The defect is slight, however, and may be ignored for most purposes.

Coloured Bodies

A natural separation of wave-lengths is made by any coloured body, for when sunlight falls on it, it scatters back to the eye only a certain selection of wave-lengths, whose resultant colour is that which we call the colour of the body. The remainder is absorbed—or, in some cases, if the body is partially transparent, transmitted. It follows that the colour of a body depends on the kind of light which illuminates it. If from that light some of the colours of the complete visible spectrum are absent, and these include colours which the body normally scatters, its colour may appear quite abnormal. This may even be so when the colours of the illuminating light, though all present, are in different proportion from those in sunlight. Everyone knows the difficulty in matching wools in artificial light, for example. The only way to be sure that two sources are identical in the quality of the light they emit is to obtain their spectra by the spectroscope, and

measure the amount of light at each wave-length. To achieve a desired result we have at our choice a variety of sources and of "light filters," which absorb certain wave-lengths and transmit others. By suitable combination we can usually get fairly close to the illumination we require.

Radiation and Temperature

The relative amounts of light of different wave-lengths in the continuous spectrum of a glowing solid or liquid body depend on the temperature of the body. This is shown best by curves (Fig. 65) giving the amount of energy at different wave-lengths. At each temperature there is a wave-length for which the energy is a maximum, and this wave-length is smaller the higher the temperature. For "black bodies" (see I, 185)—and most solids and liquids approximate to the condition of black bodies at high temperatures—the product of the absolute temperature T and the wave-length for maximum energy λ_m, is a constant. This is known as *Wien's Law*, and it is expressed by the equation

$$\lambda_m T = 29{\cdot}4 \times 10^6. \quad . \quad . \quad . \quad . \quad (6{\cdot}10)$$

when λ_m is measured in angstroms and T is the *absolute* temperature. Bodies only slightly warm do not radiate sufficient light to be visible, their radiation being entirely in the infra-red. As the temperature rises they become first red hot, since red light is the first part of the visible spectrum to be emitted in observable quantity; then yellow, and finally white, when the whole visible spectrum is radiated.

Fluorescence

The curves in Fig. 65 apply only to radiation resulting from heat (*i.e.* relative motion of molecules) either directly or indirectly. Light is sometimes radiated by *fluorescence*, in which there is no simple rule for the colours radiated. Fluorescence occurs when light of short wave-length falls on certain substances, such as paraffin, vaseline, etc. The individual

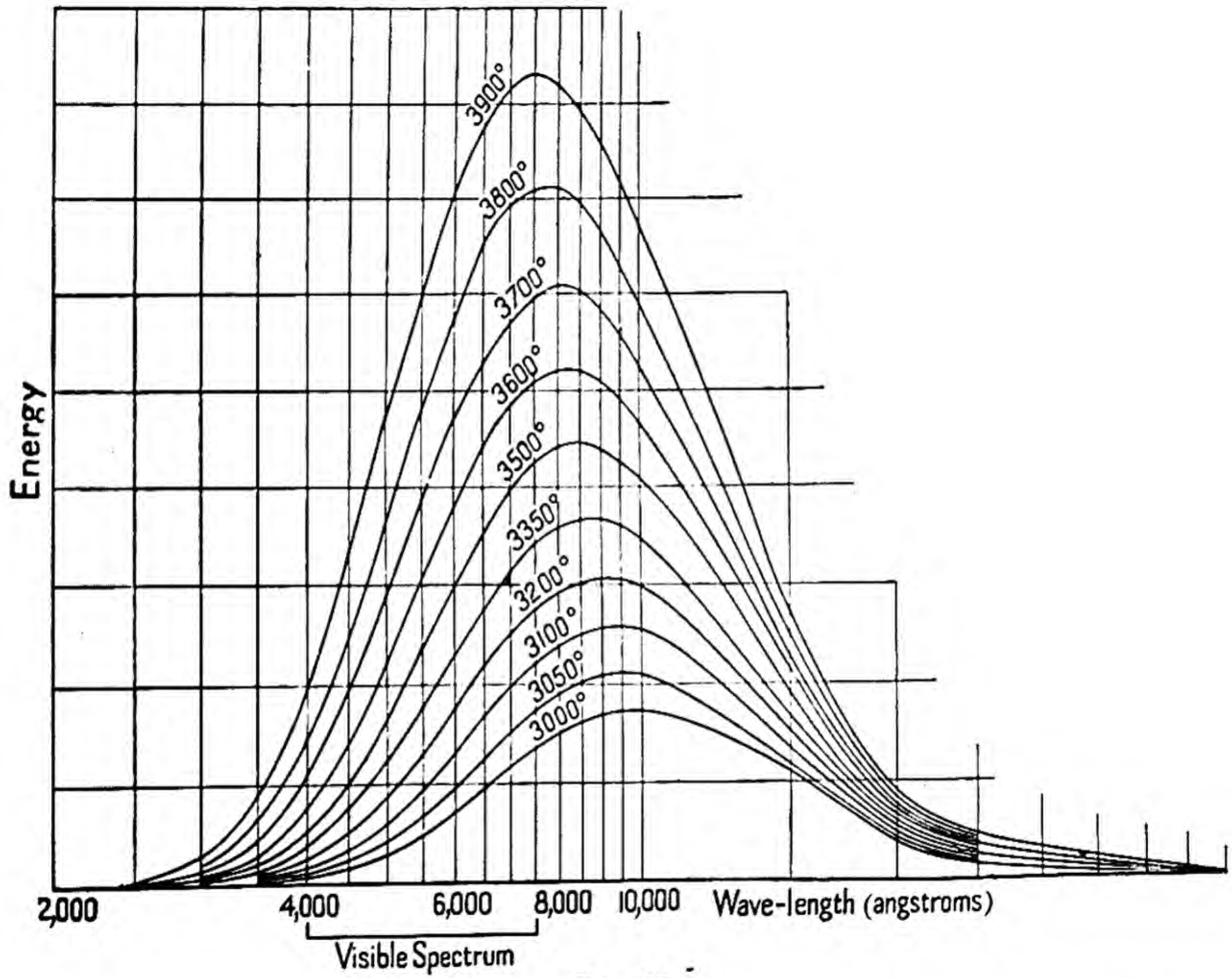


Fig. 65

Curves showing amount of energy, on an arbitrary scale, in the spectra of a black body at various temperatures on the absolute Centigrade scale

molecules of the substance absorb the energy in the incident light, and radiate it again as light of a different—usually greater—wave-length. It would take us too far afield to discuss the mechanism in the molecules by which this is achieved. When the radiation continues after the incident light has been withdrawn, the phenomenon is called *phosphorescence*. Luminous paint, such as that used on clock dials, affords a common example. Ultra-violet radiation is usually more effective than visible light in producing fluorescence or phosphorescence.

EXERCISES

1. What is meant by dispersion ? How would you separate the colours in a given beam of light, and what kind of conclusions could you draw from the character of the spectrum produced ?

2. Describe the spectroscope. What would you expect to observe in it if light from (*a*) the Sun, (*b*) a red-hot poker, (*c*) a mercury vapour lamp, were allowed to pass in turn through the slit ?

3. Explain the meaning of chromatic aberration, and state how you would construct a telescope lens in which it is reduced to a minimum.

4. Why is it that some bodies appear coloured although illuminated by white light ? How does the colour of a body depend on the character of the illumination in which it is seen ?

5. Describe and explain the changes in appearance of an iron rod when gradually heated in a dark room from the temperature of the air up to its melting-point. How are the changes in appearance associated with the changes in temperature and in the spectrum ?

6. Give an account of the processes of emission and absorption of light in terms of atomic structure.

CHAPTER VII

THE WAVE THEORY OF LIGHT

UNTIL the last chapter was reached it was of no advantage to think of light in terms of any theory at all. In the last chapter it was helpful, though not necessary, to adopt the wave theory in order to visualize the difference between light of different colours. In this chapter we shall mention a few phenomena in which a vivid picture of what is occurring can be obtained only by means of the wave theory.

The Wave-front

We think of a point source of light as the origin of a wave-train travelling outwards in all directions. If c is the velocity of light, then at a time t after radiation begins, the light has reached the surface of a sphere of radius $r = ct$. This surface is called the *wave-front*. At a considerable distance from the source, a small portion of it shows very small curvature, and may be considered as a plane : it is called a plane wave-front, and the rays which terminate there are effectively parallel. A plane wave-front may be obtained also, of course, by making the light parallel by means of a lens or mirror. The waves in the train are transverse vibrations which, in ordinary light, take place in all directions at right angles to the direction of propagation of the light. It is clear that all points in a wave-front have the same phase (see I, 199).

Propagation of Wave-train

As time goes on, of course, the wave-train advances. The manner in which it does so was viewed by Huygens in the following very useful way. Each point of the wave-front at

any instant can be regarded as the origin of a new spherical wave, and, at the next instant, the new wave-front will be made up of the fronts of these infinitely numerous secondary waves. In Fig. 66 the wave-front advancing from I to II is shown according to this idea. If we imagine an infinite number of secondary waves instead of the finite number shown, we can see that the wave-front II will be a smooth sphere concentric with I instead of the corrugated figure shown.

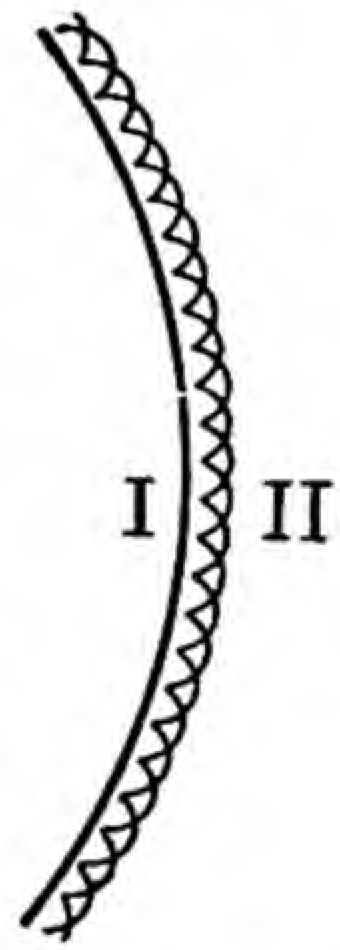


FIG. 66
Propagation of Wave-front according to Huygen's Principle
I Initial Wave-front
II Wave-front at the next instant

One difficulty which this conception suggests is that it does not explain why the wave is not propagated backwards as well as forwards, for the small spheres centred on I have portions on the left as well as on the right of I. The reason, as a mathematical investigation shows, is that the portions of the waves which would be expected on the concave side of I all neutralize one another by what is called *interference*, and produce no result. Interference has already been considered (I, 205) in connection with sound. It occurs when two or more waves simultaneously give a particle of the medium different motions. In such cases the actual motion of the particle is the resultant of the individual ones, and if it happens that the resultant vanishes, the particle does not move. It can be shown that this happens everywhere behind the wave-front I when all the secondary waves are taken into account. The wave, therefore, goes forwards but not backwards.

A similar thing occurs with a wave-front made plane by passage through a lens, as in Fig. 67, where the successive wave-fronts of the light from S, the principal focus of the lens AB, are represented by continuous lines and the rays

of light by dotted lines. The light spreads out in all directions, but that which meets the lens is thereafter confined between the rays AC and BD, and no light goes to a point such as

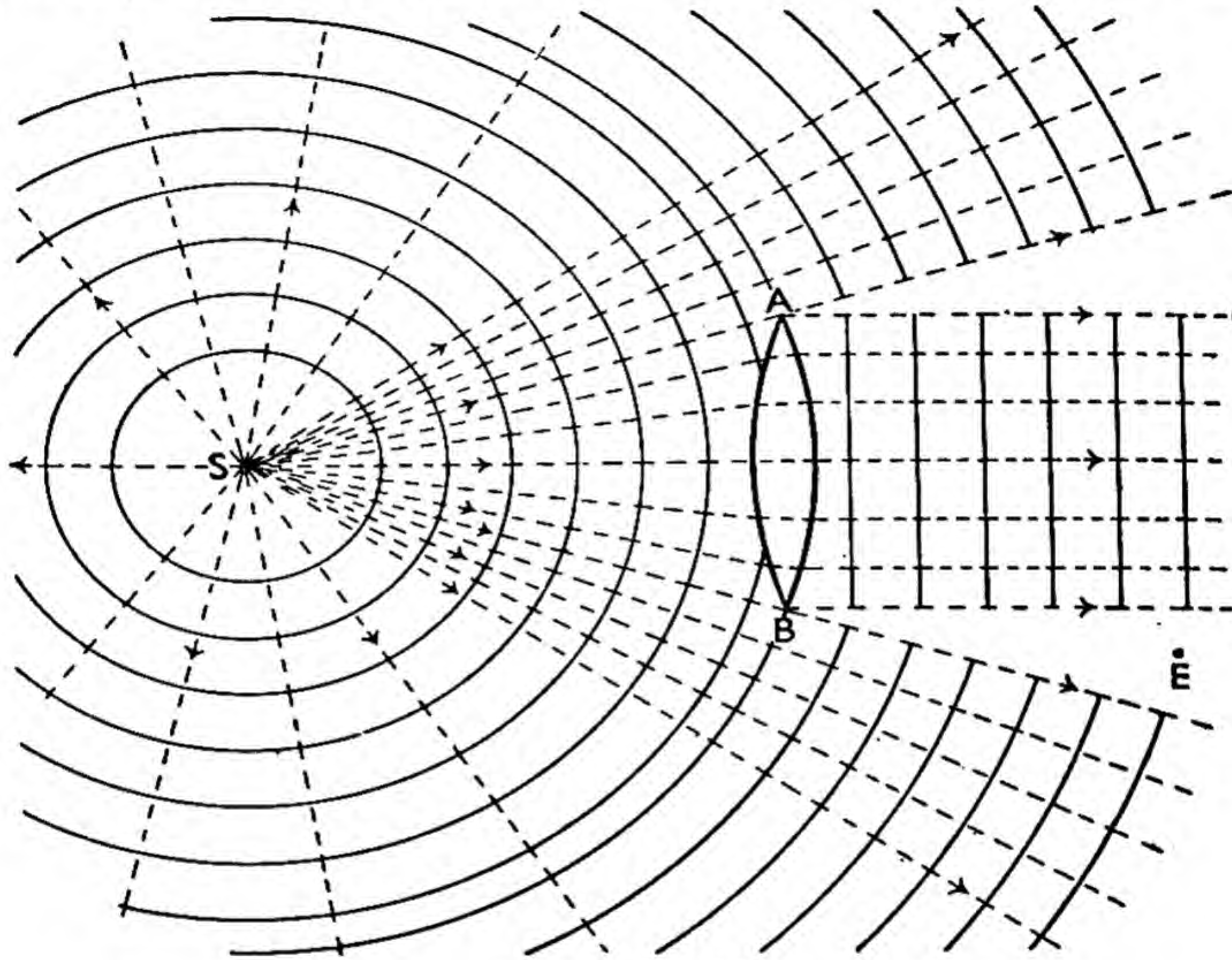


FIG. 67

Successive Wave-fronts of beam of light refracted by a lens

S Principal Focus of Lens

Continuous lines = Wave-fronts

Dotted lines = Rays

E. Again it can be shown that this is because the resultant effect there of the secondary waves from the earlier wave-fronts is zero.

The Diffraction Grating

Consider now a plane wave, restricted between two parallel lines, advancing normally on a linear set of alternate opaque

and transparent spaces—such as, for instance, a series of parallel wires with spaces between. If this structure (GH in Fig. 68) were not there the wave would go on its course, producing illumination everywhere within the boundaries ABC and DEF of the beam, and there would be darkness at all external points, such as K, because all the secondary wavelets originating in the successive wave-fronts would neutralize one another there. The result of interposing GH, however, is to cut out some of these secondary wavelets, and it does not then follow that the remainder will still neutralize one another at K. We find, in fact, that they may not.

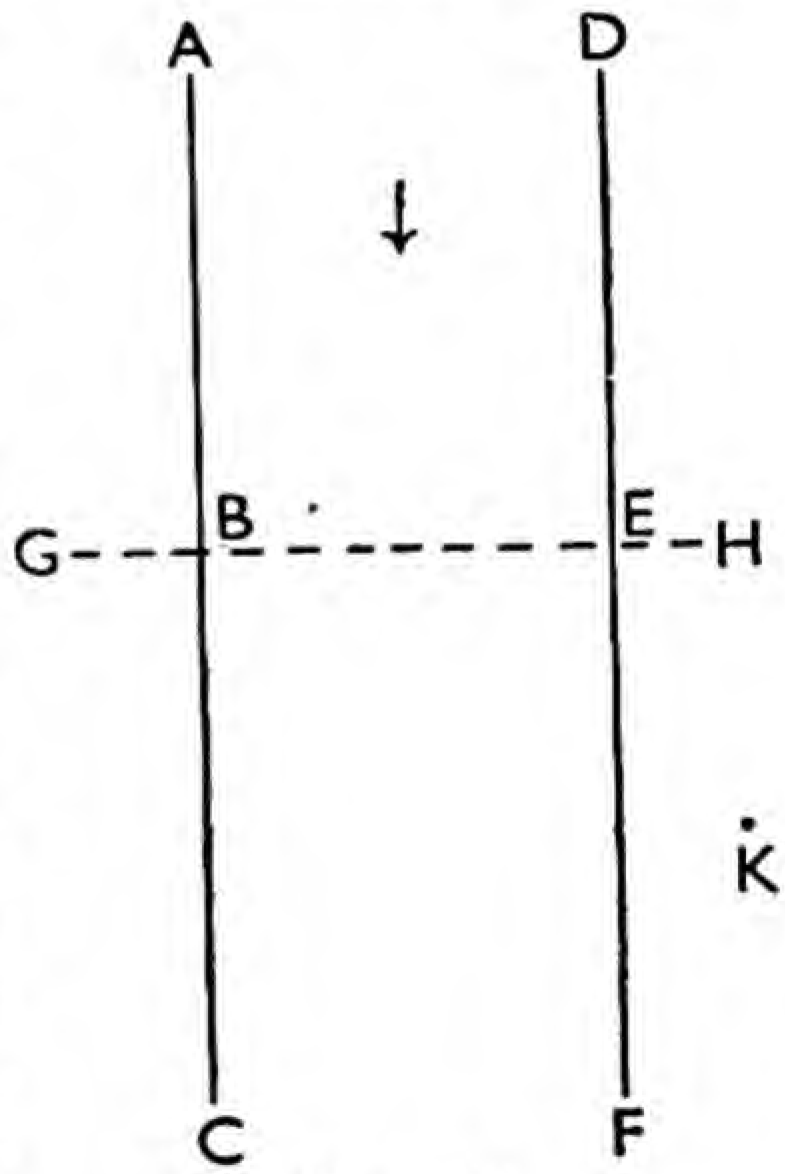


FIG. 68
Parallel Beam of Light incident normally on Diffraction Grating GH

To find the illumination, if any, at such a point as K, we proceed as follows. Consider the rays of light, AB and LM (Fig. 69, which shows a portion of Fig. 68 enlarged), which meet the interposed structure (the *grating*, as it is called) at corresponding points, B and M, of neighbouring spaces. B and M become the centres of secondary wavelets. Let us consider the advance of the wave-front in the direction BP or MQ, making an angle θ with AB or LM. If we draw MR perpendicular to BP, we see that the phase at R will be the same as that at B or M only if BR is equal to a whole wave-length or an integral number of wave-lengths. If it is, then light travelling along RP and MQ will be in the same phase at all points equidistant from R and M, and such points, lying along lines parallel to MR, will therefore be wave-fronts. Rays travelling in these directions when brought together by a lens will therefore reinforce one another, for they will both tend to move a particle of the medium in the same direction.

Furthermore, all rays from corresponding points of other spaces, such as S, T, . . . will reinforce them, for it is easily seen that rays from these points travelling in the direction θ will also be in the same phase as those from B and M in the wave-fronts parallel to MR. Hence in the direction θ there will be light.

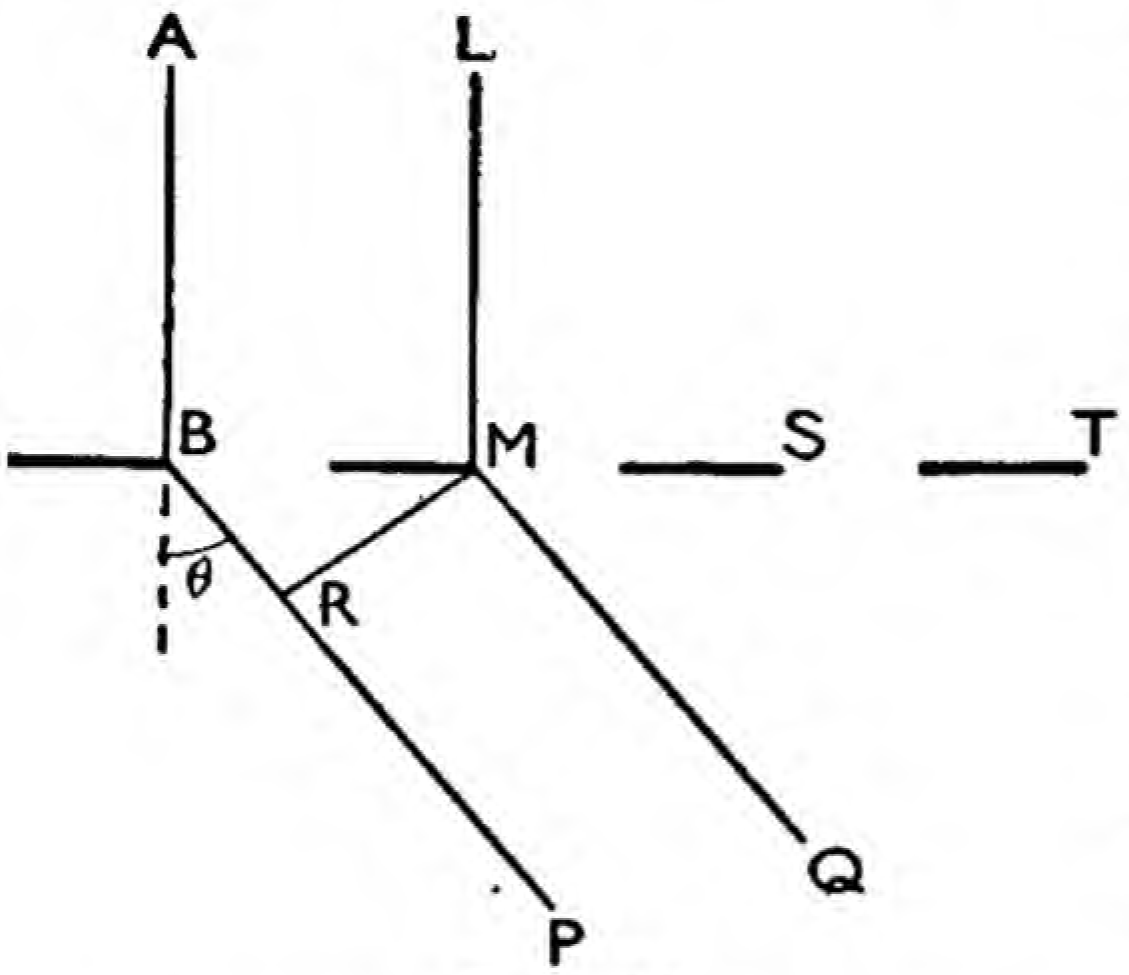


FIG. 69
Reinforcement of Waves from adjacent grating spaces in Direction θ, where BR $=$ BM $\sin\theta = m\lambda$

Spectrum Produced by Grating: The condition necessary for this to happen, as we have seen, is that BR $= m\lambda$, where m is an integer and λ is the wave-length of the light. But BR$=$BM $\sin\theta$, and if we write BM $= d$ (generally called the *grating space*) we have, as the condition for illumination,

$$d \sin\theta = m\lambda \qquad . \quad . \quad . \quad . \quad . \quad . \quad (7\cdot1)$$

Hence, for a grating of given grating space, light will be seen in the various directions θ satisfying this equation when m has successive integral values, 0, 1, 2, 3, . . .

Equation (7·1) obviously gives different values of θ for different values of λ. Hence, if we illuminate the grating with white light we shall see a spectrum, for the light passing through will be divided into the component wave-lengths which will emerge in different directions. We shall, in fact, see a series of spectra, one for each integral value of m. They are called spectra of the 1st, 2nd . . . orders when $m = 1$, 2, . . . (When $m = 0$, $\theta = 0$ for all values of λ. Hence some light continues on its original course, and is not analysed into a spectrum.) There are two spectra of each order, for one occurs on each side of the direction of the incident light.

A grating such as this (it is called a *diffraction grating*) is a very useful means of producing a spectrum, superior in some respects (but not in all) to a prism.

Diffraction

The term "diffraction" is a general one referring to the bending of light round corners in any circumstances at all. It can be shown, in fact, that whenever light meets an obstacle a slight bending occurs, whether there is a series of gaps as in a grating or not, though a grating is necessary to give a good spectrum. Light is therefore not propagated exactly in straight lines in all circumstances, but in the slight deviation therefrom which the wave-form implies. The shadow of a straight edge, for instance, shows a slight encroachment of the light into the geometrical shadow.

Polarization

It was said above that in ordinary light the motion of the particles is in all directions in the wave-front, *i.e.* in all directions perpendicular to the direction of propagation of the waves. It is possible, however, by passing the light through certain crystals (tourmaline is a good example), to restrict the movements to a single direction. For example, if the light is travelling from left to right in the plane of this page, the vibrations of the particles might also be in the plane of the page, from top to bottom and back, and not in any other direction. Light so restricted is said to be *polarized*. Rather unfortunately, it is said to be polarized in the plane perpendicular to that in which the vibrations take place. Thus, in the example given, the light is polarized in a plane perpendicular to that of the paper. Investigation is needed to determine whether light is polarized or not; the eye cannot distinguish polarized from unpolarized light.

Light can be polarized in other ways, *e.g.* by reflection. When a beam of ordinary light is incident on a plane glass surface, say, preferably at an angle of about 56°, the vibrations

perpendicular to the plane of incidence are reflected more strongly than the others, while those in the plane of incidence are mostly transmitted. By a few such reflections the light can be almost completely polarized. The best angle of incidence for the purpose (56° for glass, as just mentioned) is that whose tangent is equal to the refractive index of the reflecting substance. This is known as *Brewster's law.*

Double Refraction

When ordinary, unpolarized light passes through certain crystals (*e.g.* Iceland spar) it is divided into two refracted beams which proceed along different directions and are polarized in planes perpendicular to one another. The two refracted rays for a single incident ray are known as the *ordinary* and *extraordinary* ray. The former obeys Snell's law, but the latter does not, the angle of refraction differing for the same angle of incidence if the *plane* of incidence changes.

Nicol's Prism: The separation of the two refracted rays is brought about most effectively by a *Nicol's prism,* which consists of a crystal of Iceland spar cut in a certain manner into two parts which are cemented together by a film of Canada balsam. This substance has a refractive index lying between the refractive indices of Iceland spar for the ordinary and extraordinary rays. The angle of incidence is therefore arranged so that the ordinary ray is totally reflected internally on passing from the spar to the balsam, while the extraordinary ray goes through.

If a beam of *polarized* light falls on a crystal which has the polarizing property, it is transmitted for certain positions of the crystal, but as the latter is rotated there comes a position for which the light will no longer pass. Further rotation restores the passage of light, and the amount transmitted increases up to a maximum at 90° from the opaque position, and then falls off again, reaching another zero value at 180° from the first. For example, if unpolarized light is incident

on a Nicol prism, the transmitted light is polarized, and if this falls on another Nicol prism it may pass through in whole or in part or not at all. In the last-named event the Nicols are said to be *crossed*. If the second Nicol is rotated through an angle α from the crossed position, the intensity of the light which passes through it is $I \sin^2 \alpha$, where I is the intensity of the light which falls on it. This affords a useful means of reducing the intensity of a beam of light in an accurately known proportion if we do not mind the light being polarized.

EXERCISES

1. Give an account of the wave theory of light, explaining how the waves are propagated in space from one wave-front to another.

2. What is a diffraction grating? Explain why, although it may consist only of a set of obstacles, it transmits light in directions in which light would not travel if the grating were removed.

A grating containing 6,000 spaces per cm. is illuminated normally by light of wave-length 5,000 angstroms (1 angstrom $= 10^{-8}$ cm.). Calculate the angles at which the light will be transmitted.

3. Explain the difference between polarized and unpolarized light. How would you determine whether a beam of light was polarized or not?

4. Describe the phenomenon of double refraction, and explain how it may be used to obtain a beam of polarized light.

PART III

ELECTRICITY AND MAGNETISM

CHAPTER VIII

STATIC ELECTRICITY

In the opening chapter we gave an account of the electrification of bodies in terms of atomic structure. The atoms, normally containing equal numbers of electrons and protons (unit negative and positive charges respectively), are broken up, and electrons pass from one body to another, leaving the former positively and the latter negatively charged. This is not the normal condition of matter, and at the first opportunity the positively charged body acquires electrons and the negatively charged body expels electrons, so as to recover the neutral state.

Conductors and Non-conductors

The ease with which this is done depends on the atomic constitution of the body. In some substances electrons move fairly easily, while in others they find movement difficult. This difference is expressed by what is called the *electrical conductivity* of the body. Substances through which electrons move easily are called good conductors, and others bad conductors. Generally speaking, among solids metals are good conductors and non-metals bad conductors. If materials are arranged in the order of their conductivity, it is found that although there is no sudden transition from a group of very good to a group of very bad conductors, there is yet

a reasonably clear dividing line, so that we can speak with definiteness of *conductors* on one hand and *non-conductors*, or *insulators*, on the other. It must be understood, however, that even the worst conductor allows electrons to pass through it to some extent.

It is only insulators that can be electrified by friction in the manner already described, unless special precautions are taken. The reason is that the human body is a moderately good conductor, and if a conducting rod be held in the hand and rubbed, the atoms are restored to their normal state as fast as they are broken up, by the passage of electrons from the rod to the Earth or the Earth to the rod, as the case may require. (The Earth must be regarded as containing free electrons, and as being able to accommodate many more, without being observably electrified, owing to its great size. Any electrified body, whether charged positively or negatively, immediately becomes neutral when connected with the Earth either directly or through a conductor.) It appears to be always the electrons that move, and not the positively charged atoms (or "ions," as they are called). This would be expected, because of the much smaller mass of the electrons. If, however, a conductor be held by an insulating handle, so that electrons cannot pass between it and the hand, it also can be electrified by friction. In all experiments on frictional electricity the apparatus used must be quite dry, otherwise any electrification produced is destroyed, since moisture has conducting properties.

Induced Charges

Suppose, now, we have a conductor on an insulating stand, so that it cannot lose any charge that it may acquire (dry air at ordinary pressures is insulating). We have seen that if a charged rod is brought near it a charge of the opposite kind is "induced" in the conductor. To take a particular case, if the approaching rod is negatively charged the conductor acquires a positive charge. But this means

that it has lost electrons, and we have seen that electrons cannot escape from it. The explanation is that the atoms are indeed broken up, and the electrons are repelled to the farther side of the conductor, the side nearest the rod thus containing a preponderance of positively charged ions. The conductor is therefore charged oppositely in different parts (Fig. 70). If it is now touched by the finger, the electrons on the far side escape to the Earth, and if first the finger and then the rod be removed, the conductor is left with an excess of positive charge which spreads all over it, so that it is positively charged as a whole.

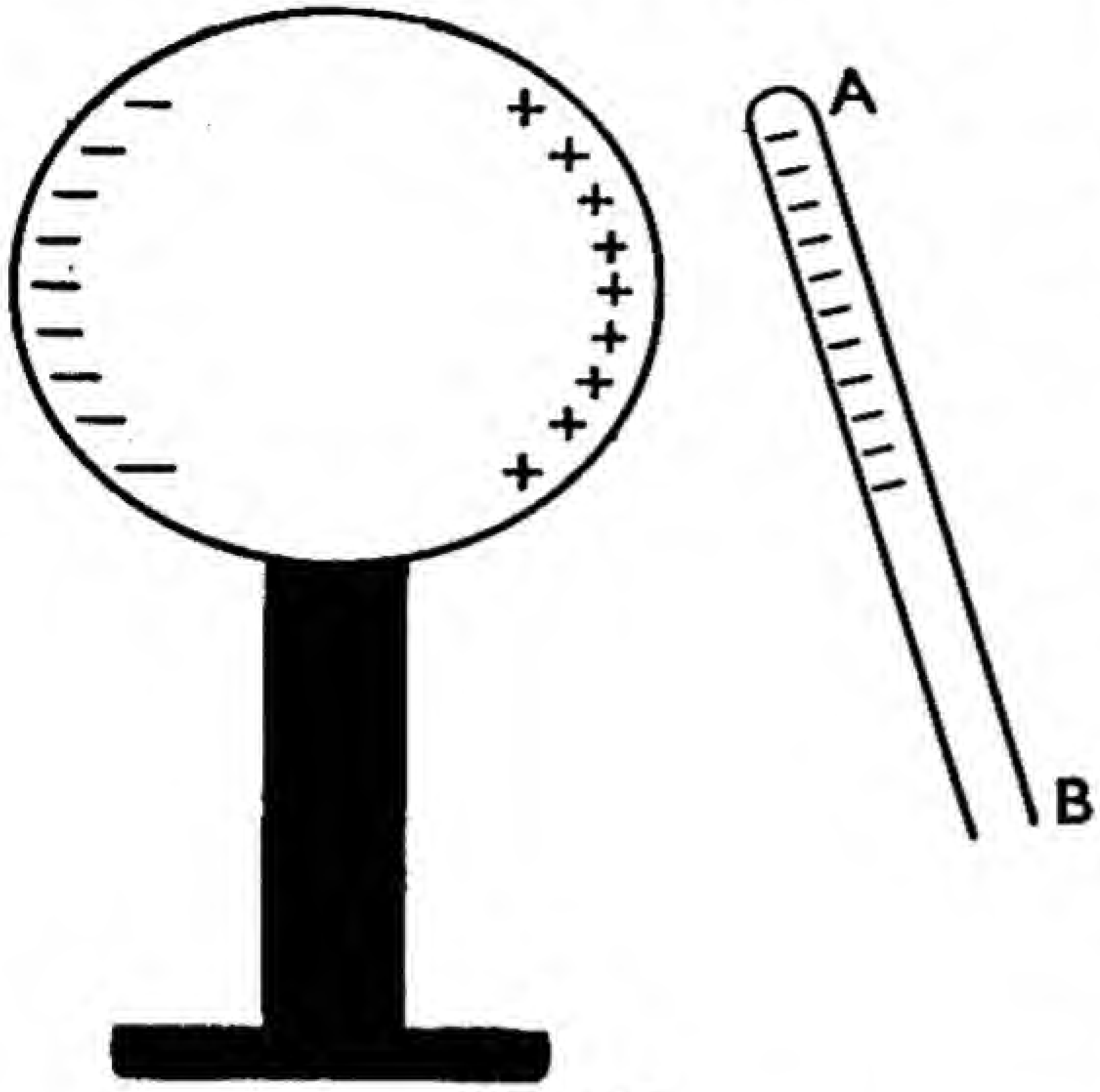


FIG. 70
Charging of Conducting Sphere by Induction from Charged Rod AB

The Proof-plane and the Electroscope

The study of the electrification of bodies is facilitated by two simple pieces of apparatus known as a *proof-plane* and an *electroscope* respectively. The former consists of a small conducting disc (*e.g.* a penny) at the end of an insulating handle (of sealing wax, for example), as in Fig. 71. If the disc is placed in contact with a charged body it becomes effectively a part of the body, and the charge is shared between the disc and the body. The disc when removed, therefore, has the

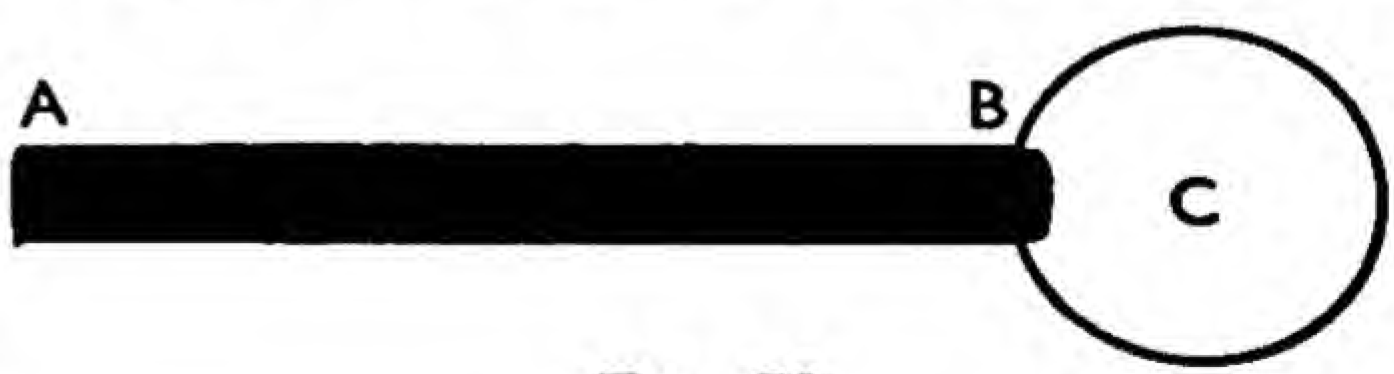


FIG. 71
Proof-plane
AB Insulating Handle
C Conducting Disc

same kind of charge as the body. This charge can then be imparted to another conductor in the same manner.

The electroscope is a very sensitive device for detecting the existence of a charge. It consists (Fig. 72) of two very thin gold leaves hanging side by side from a conducting rod surmounted by a conducting disc. An insulating vessel surrounds the leaves, and a scale placed behind them indicates the angle at which they may be separated. If a charge (either positive or negative) be imparted to the disc, it spreads to the leaves which, being similarly charged, repel one another. The greater the charge the greater the repulsion, and the leaves therefore come to rest, making an angle with one another which depends on the amount of charge given to the instrument. If, for example, the conductor in Fig. 70, after being positively charged, is touched with a proof-plane, and the proof-plane is then brought into contact with the disc of the electroscope, the leaves will diverge, both being positively charged. If a negative charge be then brought up in a similar way, the divergence of the leaves is reduced or brought to nothing ; or, if the negative charge is greater than the original positive charge, the leaves after collapsing diverge again, this time through the mutual repulsion of their *negative* charges. The electroscope can, of course, be charged by induction in the same manner as the conductor mentioned above.

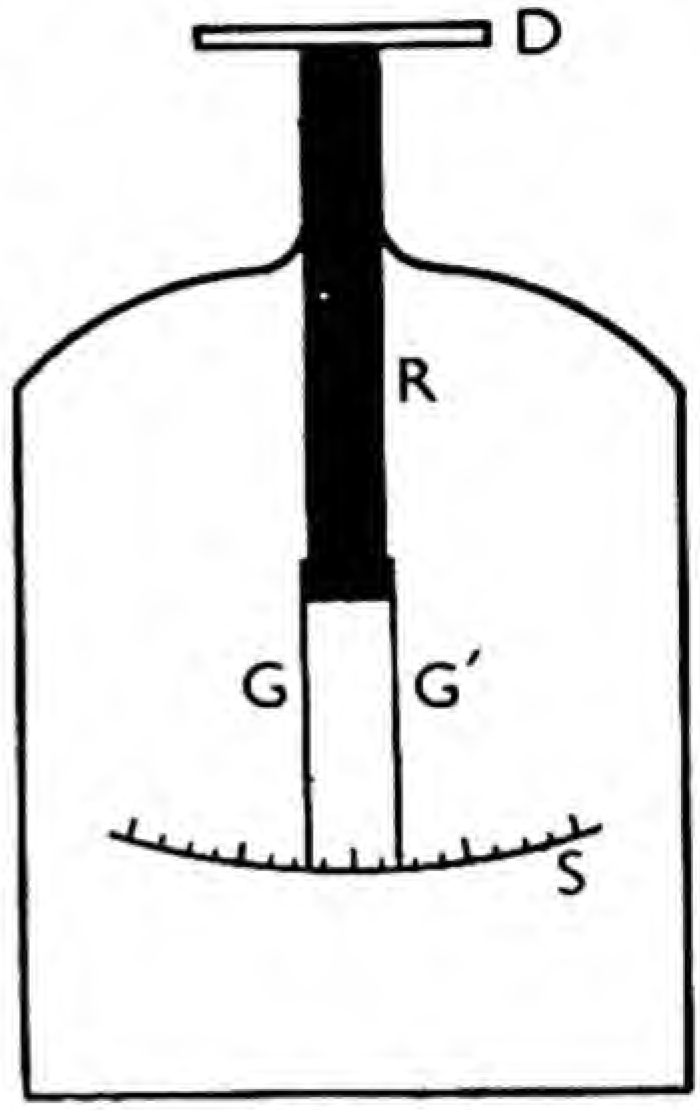


FIG. 72
Gold-leaf Electroscope
GG′ Gold Leaves
D Conducting Disc
R Conducting Rod
S Scale
The Instrument as shown is uncharged

Distribution of Charge on Surfaces

When the electrification of bodies is examined by means of these instruments, two things are noticed. First, the distribu-

tion of charge on the surface of an electrified body is such that there is more where the curvature of the surface is great

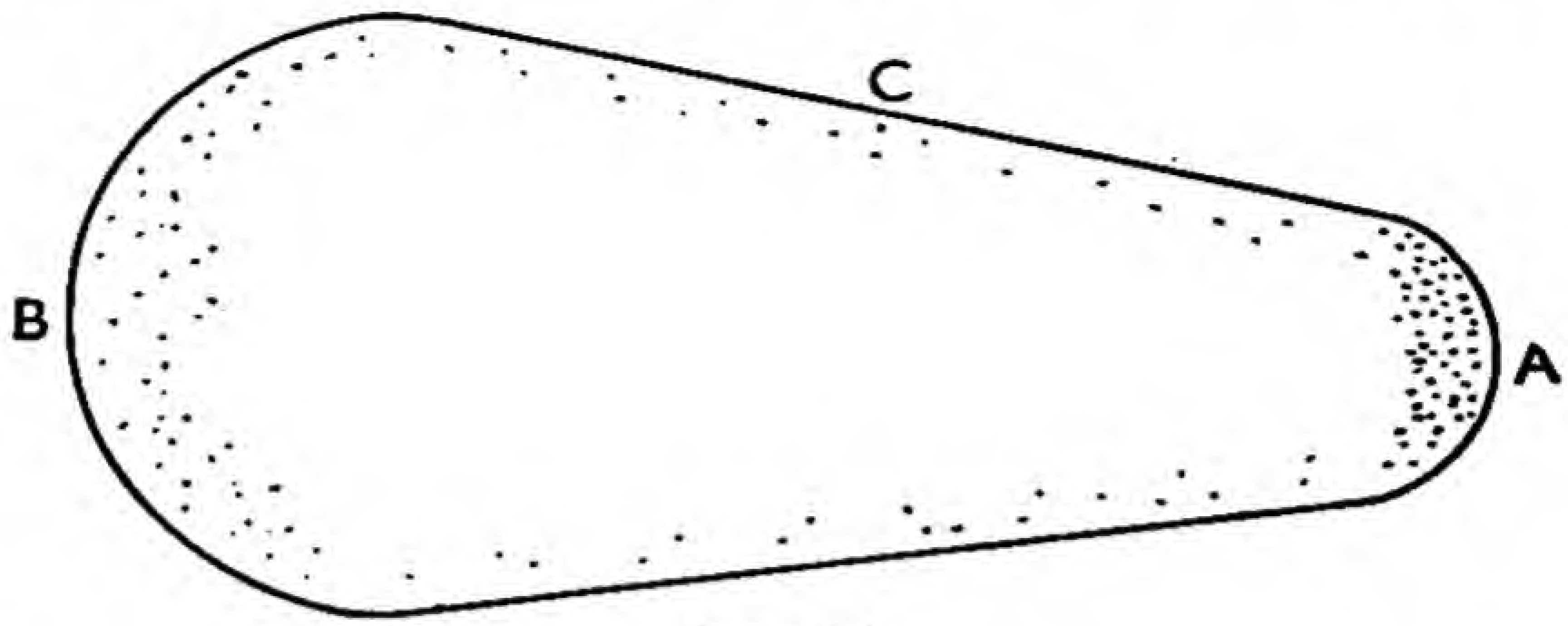


Fig. 73

Distribution of Electricity over Surface of Pear-shaped Conductor

than where it is small. On a sphere the charge is uniformly distributed, but on a figure such as that shown in Fig. 73 the charge is greater at A than at B, and greater at B than at C. This may be shown by placing identical proof-planes in contact with the surface at the three places, and comparing the amounts by which they make the leaves of an electroscope diverge. If A were a sharp point, the density of charge there would be still greater.

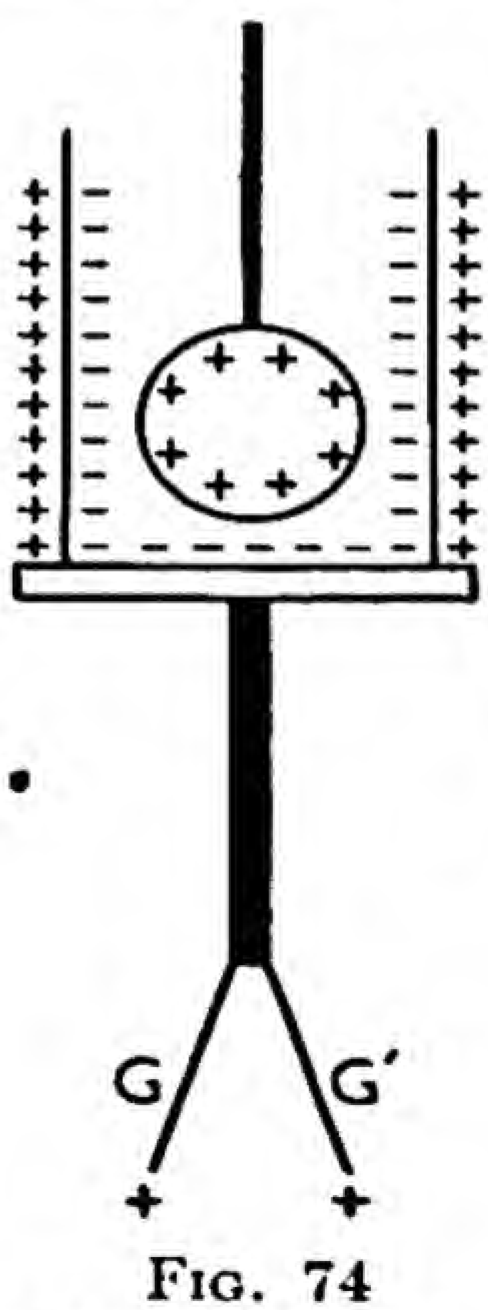


Fig. 74

Faraday's Ice-pail Experiment GG′ Gold leaves of Electroscope

The second fact is that when a hollow conductor is charged it is only the outer surface that is electrified. This again may be proved by the proof-plane and electroscope, and it plays an important part in the explanation of Faraday's famous "ice-pail" experiment. A metal (conducting) can is placed on the disc of an electroscope, and an insulated charged conductor (positively charged, say) is held inside (Fig. 74). This attracts electrons to the inner surface, leaving the outer surface and the electroscope positively charged. The

leaves therefore diverge. If, now, the conductor is made to touch the inside of the can, its positive charge neutralizes the inner negative charge, but the divergence of the leaves is not changed, for that depends on the unaffected outer positive charge. It remains even when the conductor has been removed, the conductor itself being then found to be uncharged. The charge has been transferred entirely to the outside of the can and the electroscope.

Coulomb's Law

It has already been said that the force of electrical attraction and repulsion is much greater than that of gravitation under ordinary laboratory conditions. It resembles gravitation, however, in that it is inversely proportional to the square of the distance between the charged bodies. This can be proved approximately by direct experiment, but the chief evidence for the law is the universal agreement of experiments with rather remote deductions which depend on its truth. It is known as *Coulomb's law*.

The Electrostatic Unit of Charge

It is in terms of this law that a unit of measurement of electric charge has been chosen. The natural unit would, of course, be the charge of an electron or proton, but these particles had not been discovered when the foundations of electrostatics were laid, and in any case their charges are far too small to be convenient units for measurement. The unit actually chosen—*the electrostatic unit of quantity of electricity*, or *E.S.U.*, as it is called, is the amount of charge which, when placed one centimetre away from an equal and similar charge, in a vacuum, or, for practical purposes, in air,* repels it with a force of 1 dyne. It is then a fact of experiment that a unit positive charge attracts a unit negative charge with a force of 1 dyne when they are 1 centimetre apart. The charge on a proton or electron is about 5×10^{-10} E.S.U.

* The reason for specifying the medium will be understood later

The General Inverse Square Law

It follows from Coulomb's law that if two unit charges are separated by a distance r, the force between them is $\frac{1}{r^2}$. Further, if one of the charges, instead of being a unit, is E units, the force will be E times as great, namely, $\frac{E}{r^2}$, since each of the E units will exert a force of $\frac{1}{r^2}$. If, again, the second charge is increased to e units, the force will be increased to $\frac{Ee}{r^2}$, for the same reason. We thus obtain the general equation for the force F between two charges, E and e, distant r cm. from one another, viz. :

$$F = \frac{Ee}{r^2} \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8{\cdot}1)$$

If the charges have the same sign the force is a repulsion, and if they have opposite signs it is an attraction.

The Electric Field

A charged body is thus the origin of a force which acts on other charged bodies which happen to be in its neighbourhood. If they are very close to the body the force is great ; if they are distant it is small. The whole space surrounding a charged body is called its *electric field*. It is described by what is called the *intensity* (or sometimes the *strength*) of the field, which is the resultant force, in magnitude and direction, which would act on a *unit* positive charge placed in the field. Obviously the intensity varies from place to place, being greater near the charged body than at a distance.

Lines of Force : If there were only one charge in the world it would be easy to describe its field, for the resultant force at any point would be directed towards (or away from) the charge, and would be equal to the magnitude of the charge divided by the square of the distance of the point from it. The intensity would thus fall off rapidly in all

directions. Consider, however, the field due to two charges, E and e. To find the intensity at any point we must find the resultant of the forces on a unit positive pole placed at that point, arising from the two charges. Thus, in Fig. 75, if we suppose E to be a negative charge and e a positive charge, the intensity at A would be represented by a vector such as AB, the diagonal of the parallelogram of forces of which the sides are respectively proportional to $\frac{E}{r_1{}^2}$ and $\frac{e}{r_2{}^2}$. By mapping out the whole field in this way we arrive at the distribution of intensity represented in Fig. 76, which shows,

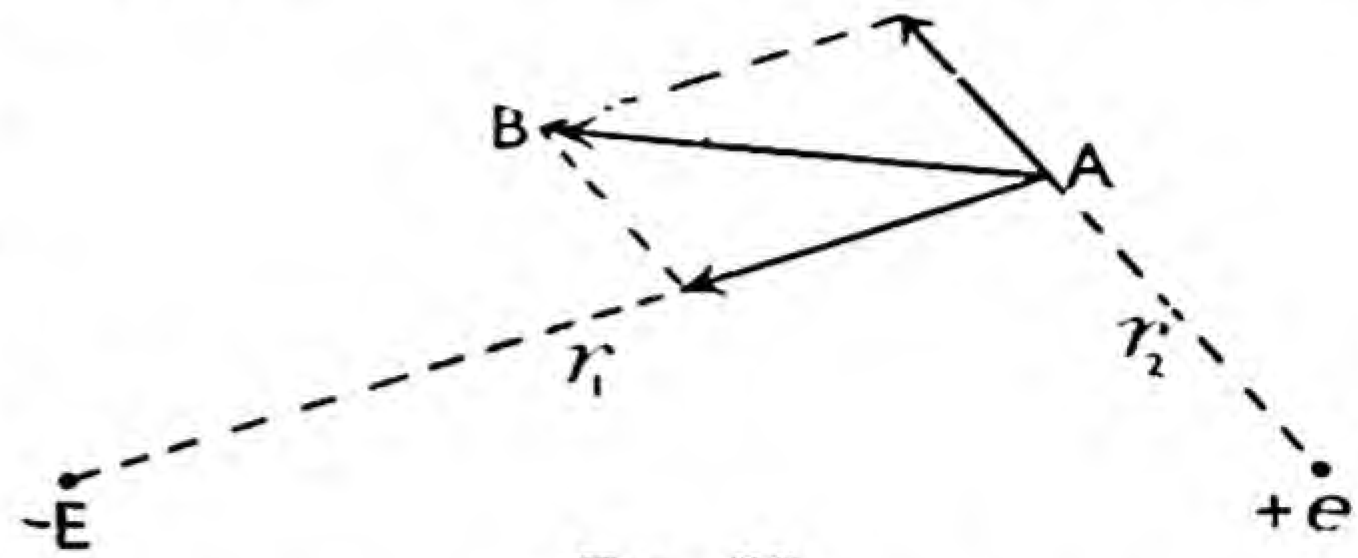


FIG. 75
Resultant Intensity AB of Field at A arising from Charges $-E$ and $+e$

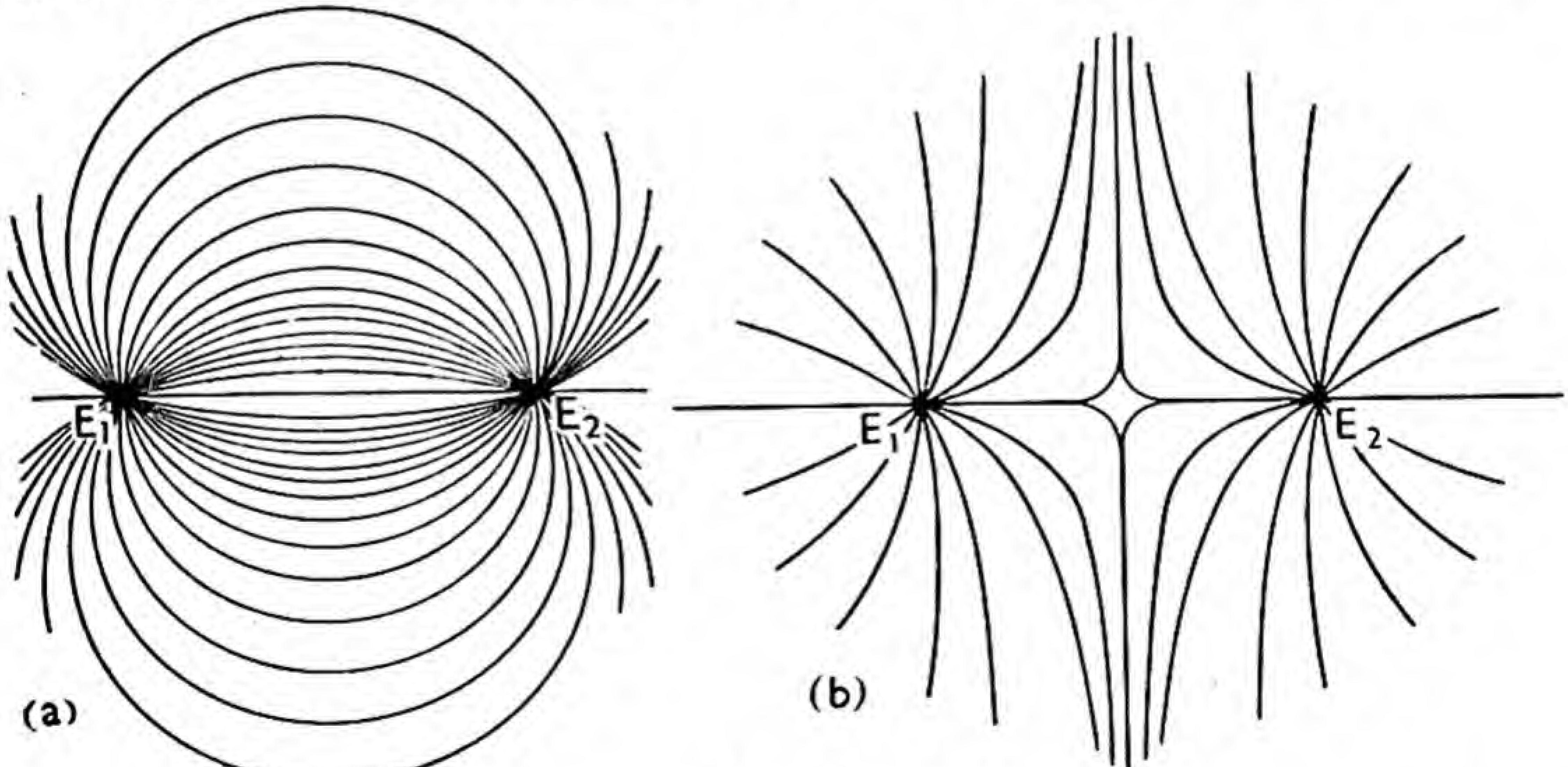


FIG. 76
Lines of Force indicating Distribution of Intensity in Field due to two Electric Charges
(*a*) Equal Dissimilar Charges at E_1 and E_2
(*b*) Equal Similar Charges at E_1 and E_2

by the directions of the lines, the direction of the force at any point in the neighbourhood of the charges when they are (*a*) equal and dissimilar, and (*b*) equal and similar. When several charges are concerned it is clear that the distribution of intensity in the field can become very complex.

Fig. 76 gives no obvious indication of the *magnitude* of the intensity at any point, but only of its direction. In fact, however, the magnitude is indicated by the degree of crowding of the lines (*lines of force*, as they are called). They are close together where the field is strong and separated more widely where it is weak. This can be made into a precise indication of the intensity by drawing the lines so that the number passing through an imaginary unit area placed normally to them at any place is proportional to the intensity at that place. This method of representing the intensity in an electric field is sometimes extremely convenient and suggestive. It is, of course, purely an artificial device, for there is an intensity at *every* point in the field, in both the strong and the weak regions, whether a line is drawn through the point or not.

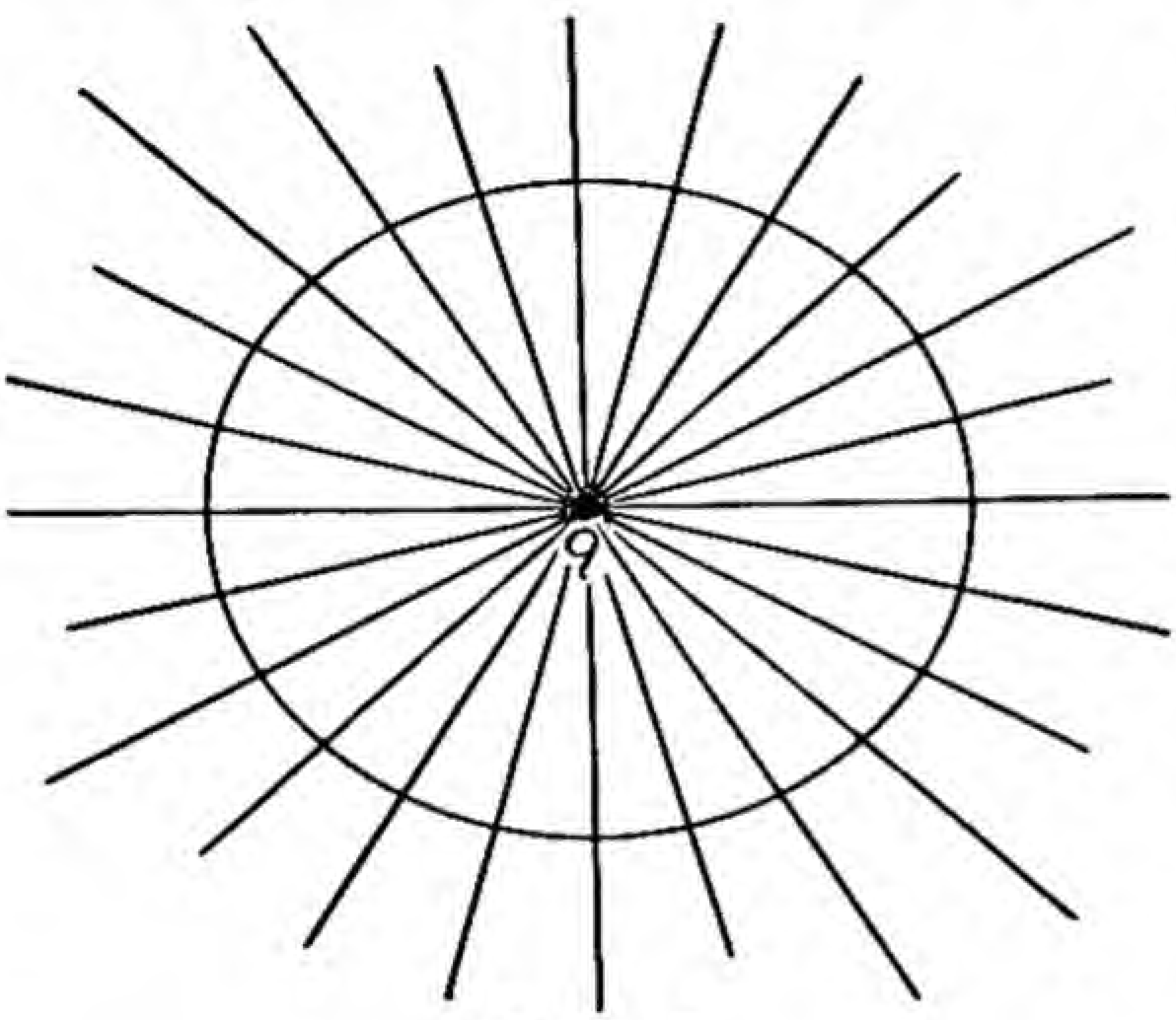


FIG. 77
Lines of Force proceeding from charge q

Let us see how these lines of force are distributed in the field of a single unit charge. Draw an imaginary sphere of unit radius round the charge, q (Fig. 77). Then the intensity at any point on this sphere $\left(\frac{q}{r^2}\text{ in the general case}\right)$ is unity, for both q and r are 1. Hence we must draw one line of force through each unit of area. The whole area of the sphere is

4π square centimetres; hence 4π lines must pass through it normally from q. We may thus say that each unit charge is the origin, so to speak, of 4π lines of force, and clearly, by the same argument, a charge of E units is the origin of $4\pi E$ lines of force. (The fact that 4π is not a whole number accentuates the artificiality of this representation, but it is none the less useful for that.) If another charge be brought into the field, the course of the lines is diverted, and they approach or recede from one another, but the number which issue from each charge remains constant. 4π lines proceed from each unit charge in all circumstances.

It has been thought by some physicists that these lines of force are not simply a convenient graphical device for representing an electric field, but indicate, though somewhat symbolically, actually existing lines of strain in the ether filling all space. This idea is not now generally held, but it has contributed very greatly to the progress of the science of electricity.

EXERCISES

1. Describe the gold-leaf electroscope, and explain how it may be given a permanent charge. How would you use it to determine whether the charge on a given body was positive or negative?

2. State Coulomb's law. Find the force between two particles having charges of $+50$ and -25 units, respectively, when they are 10 cm. apart. If they are brought into contact with one another and separated again by the same distance as before, what force then exists between them?

3. What is meant by the intensity of an electric field? Find, in magnitude and direction, the intensity of the field due to the charges mentioned in the last question, at a point midway between them, both before and after contact.

4. Explain the conception of lines of force as a means of describing the state of an electric field. How can the conception be extended to represent the field quantitatively as well as qualitatively?

5. Draw as accurately as possible the lines of force in the field of the two charges mentioned in question 2, both before and after contact.

CHAPTER IX

POTENTIAL AND CAPACITY

Potential

We have seen that electrons are able to move with considerable ease through some bodies—conductors—and the question arises: what determines whether electrons will flow, and, if they will, in what direction? Often, of course, the answer seems to be simple, as when we try to charge a conductor, held in the hand, by friction. As fast as electrons are detached, fresh electrons move from the Earth to make the conductor neutral again. We might conclude from this that electrons will flow whenever there is a positive charge for them to neutralize. But this is not so, for in other cases (Fig. 70, for instance) we have different parts of a conductor oppositely charged, but there is no flow of electrons to restore neutrality. This is accounted for in the case illustrated in Fig. 70 by the fact that the electrons are not only attracted by the positive charge on the sphere, but also repelled by the negative charge on the neighbouring rod, and the figure represents the equilibrium condition; but it is convenient to have a general name for the state of an electrified system which determines whether electrons will flow in it or not, and the name chosen is *potential*. Electricity will flow from one part of a conducting system to another if there is a *difference of potential* between the parts, but not otherwise.

The idea of potential is in many respects similar to that of temperature. Just as a difference of temperature indicates that heat will flow along a thermal conductor, so a difference of potential indicates that electrons will flow along an electrical conductor. We have no simple picture of potential corres-

ponding to the average kinetic energy of molecular motions which represents temperature, and so the idea is more difficult to grasp. It becomes simpler, however, as we get more familiar with it.

Since electrons can move through conductors, all parts of a conductor are at the same potential in the equilibrium state; otherwise electrons will flow through it until they are. This is quite irrespective of whether the conductors are charged or not, or, if charged, how the charge is distributed over them. Potential is no more dependent on electricity distribution than temperature is on heat distribution. The sphere in Fig. 70, for example, has the same potential all over, notwithstanding that its two sides are oppositely charged. Similarly, two points in an uncharged system will be at different potentials if, when an electron is placed at one, it tends to move towards the other.

Measurement of Potential

If two bodies are at different potentials and are connected by a conductor, electrons will flow from one body to the other until the potentials are equal. This again is quite analogous to the flow of heat until temperatures are equal. We need now a measure of potential, and just as we measure temperature difference by an effect of the flow of heat which it makes possible, so we measure potential by an effect of the flow of electrons which it makes possible. The effect we choose is the performance of work. When electricity flows from one point to another, it does work or work is done upon it. For instance, if the two charges in Fig. 76 (*a*) were allowed to move together as the result of their mutual attraction, work would be done, just as when a stone falls from a height by the gravitational attraction of the Earth. In their original positions they possess "potential" energy, which can be converted into work. The amount of this work can therefore be used as a measure of their difference of potential.

The idea of potential is analogous to that of temperature

in another respect ; namely, that it is only differences of potential, and not the actual value of a particular potential, that are significant. We are, therefore, at liberty to choose an arbitrary zero of potential, just as we were at liberty to choose the temperature of melting ice as an arbitrary zero of one scale of temperature. (The idea of " absolute " temperature (I, 126) arises from a particular choice of temperature scale in terms of the properties of gases ; this has no analogue in the measurement of potential.) We choose the zero of potential as the potential at a point infinitely distant (or, in practice, at a very large distance) from all charged bodies.

Potential due to Single Charge : Let us take a simple case, to make the idea clear. Suppose we have only one charged body E in space. Then the potential at a point infinitely distant from it is zero. What is the potential at some other point A distant a from the charge E (Fig. 78)? (It may be repeated that, although there is no charge at A, and possibly no matter there, it has a potential if a charge at infinity tends to flow towards or away from A.) *The potential at A is measured by the work which must be done in bringing a unit positive charge from infinity to A.* We can calculate this work, for (taking E at first to be a positive charge), the force resisting the approaching unit positive charge at any point distant r from E is $\frac{E}{r^2}$, and when the unit charge moves against this force through a small distance dr, the work which must be done on it is $\frac{E}{r^2}dr$. The work done in moving the unit charge from infinity to the point A is therefore

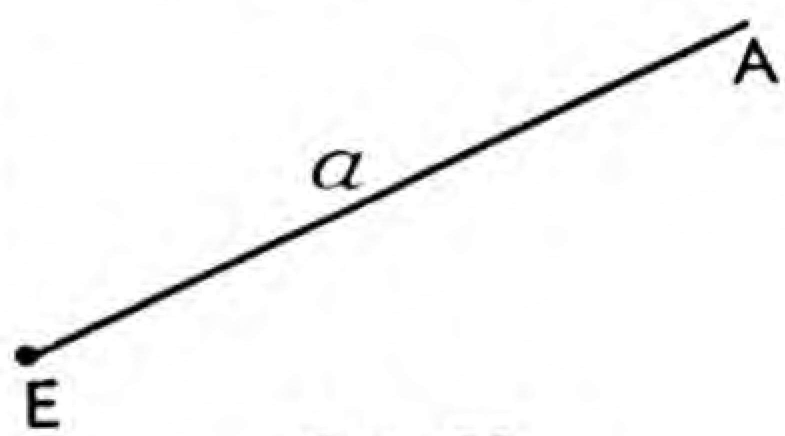


FIG. 78

Potential at A due to Charge $E = \frac{E}{a}$

$$W = \int_a^\infty \frac{E}{r^2}\,dr = \frac{E}{a} \quad . \quad . \quad . \quad . \quad . \quad (9\cdot1)$$

This quantity is the potential difference between infinity and A, and since the potential at infinity is chosen as zero, $\frac{E}{a}$ is the actual potential at A. If E had been a negative charge, work would have been done *by* the unit positive charge, for it would have been attracted by E, and the potential would have been negative. The fact that E itself is negative ensures that the sign of the potential will come out all right if the charge is given its proper sign in (9·1).

Equipotential Surfaces

We thus arrive at the general result that the potential at a point distant r from a single charge E is $\frac{E}{r}$. All points on an imaginary sphere surrounding E therefore have the same potential, for they have the same value of r. We may thus draw a series of *equipotential surfaces* around E, and they will be simply a series of concentric spheres. Fig. 79 shows the lines of force and the equipotential surfaces in the neighbourhood of such a charge. Clearly the lines of force and the equipotential surfaces are everywhere at right angles to one another, and a little reflection shows that this must be so not only in the field of a single charge but in every possible electric field. For there is no force acting along an equipotential surface, otherwise electricity would tend to flow over it in obedience to the force, and the surface could not then be an equipotential one. If, now, we take any point on a line of

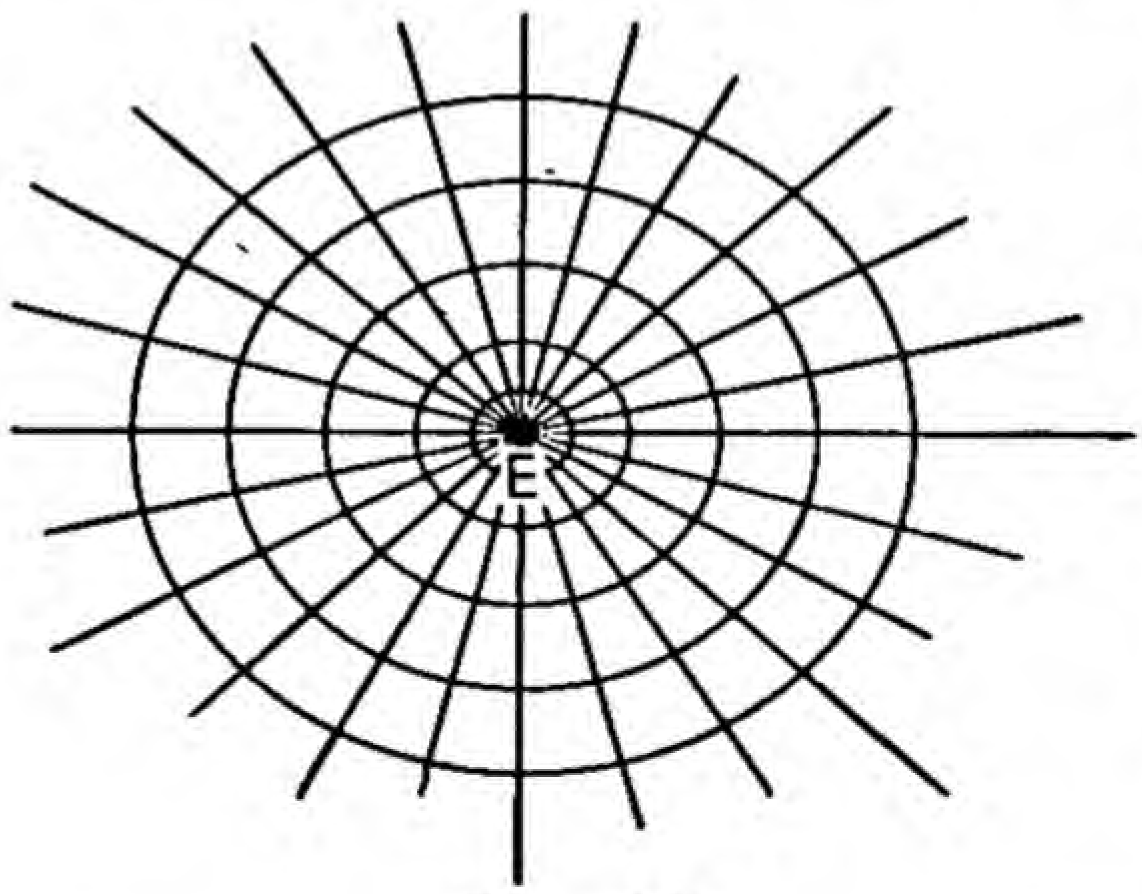


FIG. 79

Lines of Force (straight lines) and of Equal Potential (circles) in Field of Single Charge E

force, we see that the only direction through that point in which there is no force (*i.e.* in which the resolved component of the force vanishes) is the direction perpendicular to the line of force. Hence this direction must lie on the equipotential surface. An obvious but important deduction from this is that the lines of force in the field of a charged conductor meet the surface of the conductor at right angles. For the surface of the conductor must be an equipotential surface, or electricity would immediately move over it until it was so.

The potential at a point in the field of a number of charges is obtained simply by adding the potentials due to the separate charges ; in mathematical terms it is $\Sigma\frac{E}{r}$. Potential, unlike force, has no direction associated with it. The potential at a point is measured by an amount of work, and it can be shown that this work is the same by whatever path the unit charge reaches the point from infinity.

Potential of the Earth

It is often convenient in practice to take the potential of the Earth as zero. This is so nearly true that no appreciable error is caused thereby, for we know by experience that the Earth is a conductor and is practically uncharged, since any simply charged body, whether positive or negative, immediately loses its charge when "earthed." No doubt there is a slight excess or defect of electrons in the Earth at that moment (see Chap. XV), but the Earth is so vast that its charge as a whole is negligible. Hence the work done in bringing a unit positive charge up to it from infinity (*i.e.* its potential) is zero, for it exerts no force on the charge. It follows that any body which is earthed (*i.e.* connected with the Earth by a conductor) has zero potential.

Force and Potential Gradient

We may now easily obtain another important result, namely, that at any point in an electric field, the force F in any

direction is equal to *minus* the rate of variation of potential $\left(-\frac{dV}{dx}\right)$ in that direction. For consider any two neighbouring points in the field, separated by a distance dx. The potential difference dV between them is the work which must be done in taking a unit positive charge from the point of lower to that of higher potential. Now if the force on the charge in the

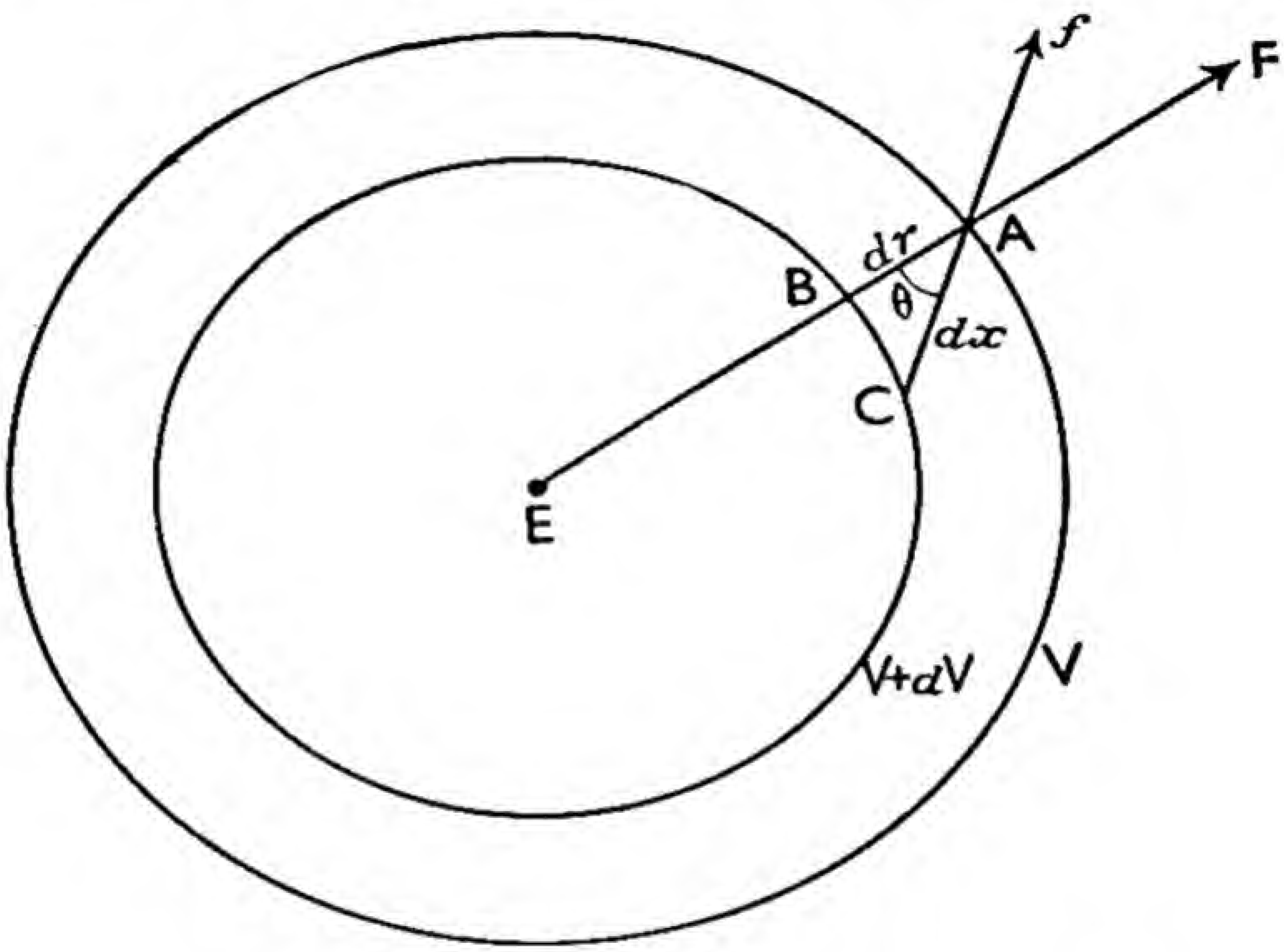


FIG. 80

Force F at A in Field of Charge $E = -\frac{dV}{dr}$

direction dx is f, this work is $-fdx$, for the force acts in the direction from higher to lower potential (just as heat flows in the direction from higher to lower temperature). Hence

$$\left.\begin{aligned} dV &= -fdx \\ \text{or } f &= -\frac{dV}{dx} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (9{\cdot}2)$$

This applies to any direction round the point. The force is greatest, of course, in a particular direction—the direction of the *intensity* of the field ; in any other direction it is simply

a component of the intensity. If the intensity be represented by F, then we have

$$F = -\frac{dV}{dr} \quad . \quad . \quad . \quad . \quad . \quad . \quad (9\cdot3)$$

where r is measured in the direction in which the potential changes most rapidly. This is illustrated in Fig. 80, where E is a positive charge, and two equipotential surfaces, of potentials V and $V + dV$, are drawn, distant dr apart. The intensity F is along the line of force, and we have $F = -\frac{dV}{dr}$, where $dr = \text{AB}$. In the direction of f, the distance, dx, corresponding to the difference of potential dV is $\text{AC} = dr \sec \theta$.

$$\text{Hence } f = -\frac{dV}{dr \sec \theta} = -\frac{dV}{dr} \cos \theta \quad . \quad . \quad . \quad (9\cdot4)$$

This agrees with the ordinary principle of resolution of forces, according to which $f = F \cos \theta$, since $F = -\frac{dV}{dr}$.

Capacity

When a body, originally uncharged, is given a charge of electricity, its potential is changed, for work is then needed to bring a unit positive charge up to the body from infinity. (If the body is negatively charged, of course, the incoming unit charge can do work by its approach, and the potential is therefore negative.) It is a matter of importance to determine by how much a given charge will change the potential, and this we can do by finding how much work is needed to bring a unit charge up to the body.

Capacity of Conducting Sphere

Consider first the simple case of a conducting sphere, and suppose it is charged with $+Q$ units of electricity. Since the curvature is the same all over the surface, this charge will

be uniformly distributed. To find the work done when an additional unit charge is brought up, we must take each small element of area on the sphere separately—for such elements are at various distances from the unit charge and therefore exert different forces on it—and add up all the resulting elementary quantities of work. The result of doing this turns out to be the same as though the whole of the charge were concentrated at the *centre* of the sphere. For example, in Fig. 81 some points on the surface are nearer to the unit charge *e*, at any particular point in its journey, than is the centre, and some points are farther away, and the effect on the whole is the same as if all points were at the distance *eO*, although, of course, all the charge is in fact on the surface. This is true whatever the radius of the sphere.

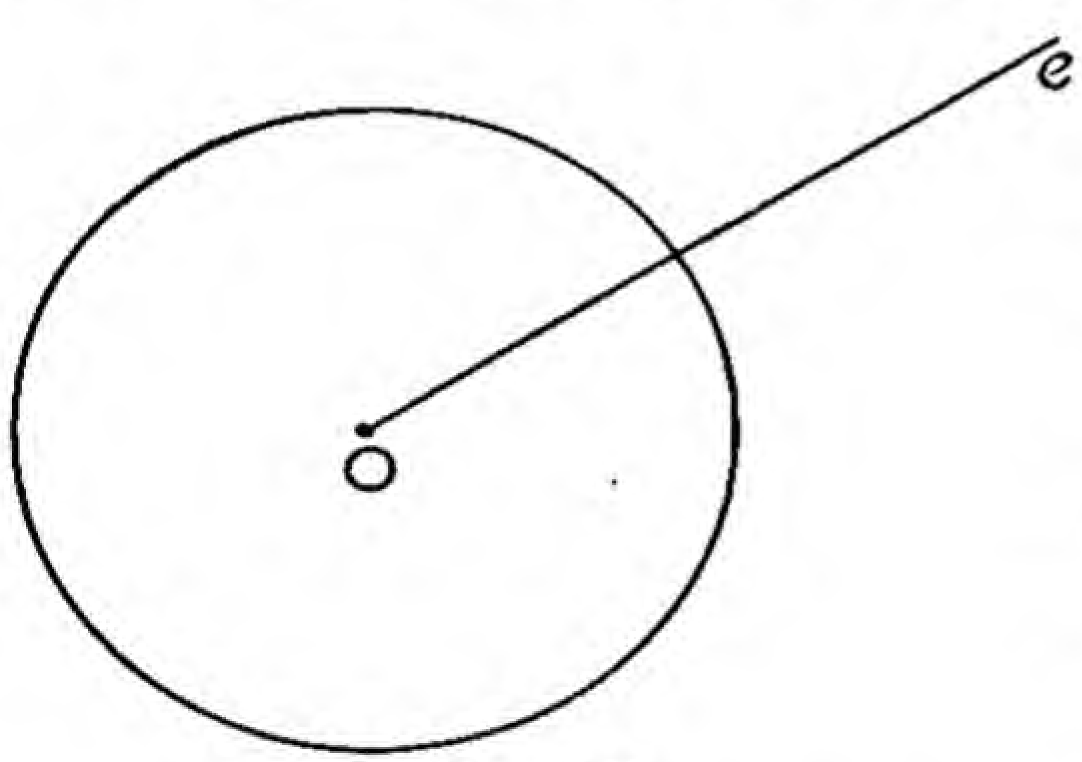


FIG. 81

Field at *e* due to uniformly charged sphere $= \frac{\text{Total charge}}{\text{Oe}}$

The force on the unit charge at an arbitrary distance r from the centre is therefore $\frac{Q}{r^2}$, and the work done in bringing the charge from infinity up to the surface is therefore $\int_a^\infty \frac{Q}{r^2}dr$, since the final position of the charge is distant a, the radius, from the centre of the sphere. This is, of course, $\frac{Q}{a}$, so that we have for the potential,

$$V = \frac{Q}{a} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9{\cdot}5)$$

It follows from this that a given charge will raise the sphere to a potential inversely proportional to the radius.

A very small sphere is raised to a high potential, and a large sphere to a low potential, by the same charge. The amount of electricity necessary to raise the potential of the sphere from zero to unity is called the *capacity* of the sphere. By putting $V = 1$ in (9·5) we see that $Q = a$; *i.e.* the capacity of a sphere is equal to its radius. The same definition applies to any conductor, whether spherical or not (though, of course, the *value* of the capacity is not the same for all conductors), and we have the general definition that *the capacity of a conductor is the amount of charge necessary to raise its potential from zero to unity.* In symbols, if Q is the charge which raises the potential to V, the capacity C is given by

$$C = \frac{Q}{V} \cdot \qquad (9{\cdot}6)$$

Following up our thermal analogy, electrical capacity clearly corresponds to heat capacity. Just as heat capacity is the amount of heat necessary to raise the temperature of a body by unity, so electrical capacity is the amount of electricity necessary to raise the potential of a conductor by unity.

Condensers

Let us now suppose there is another conductor in the neighbourhood of the sphere, which so far we have supposed to be isolated from other bodies. To take a particular case, suppose the sphere in Fig. 81 is surrounded by a larger concentric conducting sphere, of radius b, as in Fig. 82, and suppose this outer sphere is earthed. Then we know (pp. 134–35) that if there is a quantity $+Q$ of electricity on the inner

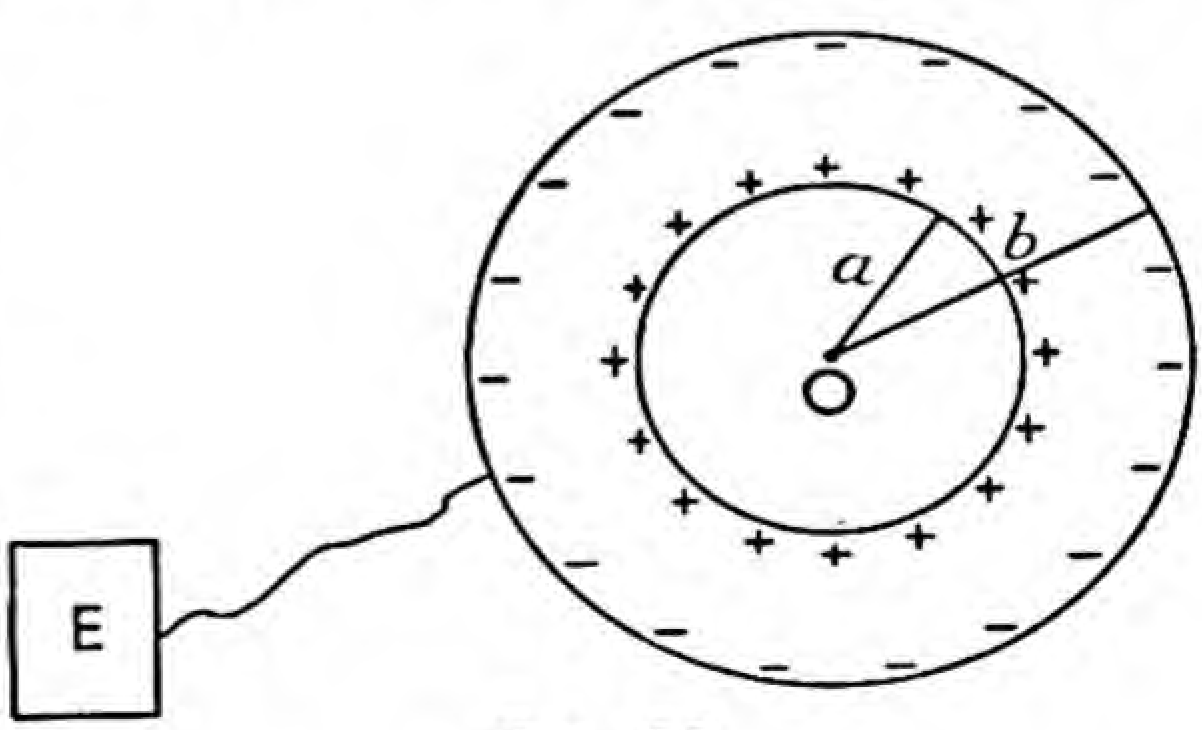


FIG. 82
Spherical Condenser
E Earth

sphere, an equal quantity $-Q$ of negative electricity will be induced on the inner surface of the outer sphere. The work done in bringing a unit positive charge up from infinity to the inner sphere is now reduced, because during most of its journey the approach of the charge is assisted by the attraction of the outer negative charge. We can at once determine the amount of the reduction without detailed calculation from the fact that the potential of the outer sphere is zero since it is earthed. The work done in bringing up the unit charge to the surface of this sphere is therefore zero, whereas previously it was $\frac{Q}{b}$, since that was the potential at the position of the surface of the outer sphere (p. 143). Hence the potential of the inner sphere is now reduced by $\frac{Q}{b}$, and is therefore

$$V = \frac{Q}{a} - \frac{Q}{b} \quad . \quad . \quad . \quad . \quad . \quad (9{\cdot}7)$$

Putting this value for V in the general equation (9·6) we therefore obtain

$$C = \frac{1}{\frac{1}{a} - \frac{1}{b}} = \frac{ab}{b-a} \quad . \quad . \quad . \quad . \quad . \quad (9{\cdot}8)$$

The effect of the outer sphere is therefore to increase the capacity of the inner sphere, for the above quantity is clearly greater than a. It is, in fact, generally true that the capacity of a conductor is increased by the neighbourhood of a second conductor, *i.e.* it takes a greater charge to raise its potential by a given amount. The two conductors together are said to form a *condenser*.

Condensers are of great importance in practical applications of electricity, and they have various forms. A *parallel-plate* condenser is a simple and common form. In this we have simply two parallel plane sheets of metal, one of which is earthed. If the area of each plate is A and the distance

between the plates is d, it can be shown that the capacity is

$$C = \frac{A}{4\pi d} \quad . \quad . \quad . \quad . \quad . \quad . \quad (9 \cdot 9)$$

so that it is increased by having large plates and putting them close together. The familiar *Leyden jar* is a condenser almost of this type (Fig. 83). It consists of a glass jar of which the greater part of the inner and outer surfaces is coated with tinfoil. A chain hanging from an insulated conducting rod rests on the inner surface, and the charge is applied to this surface through the rod. The outer surface is earthed.

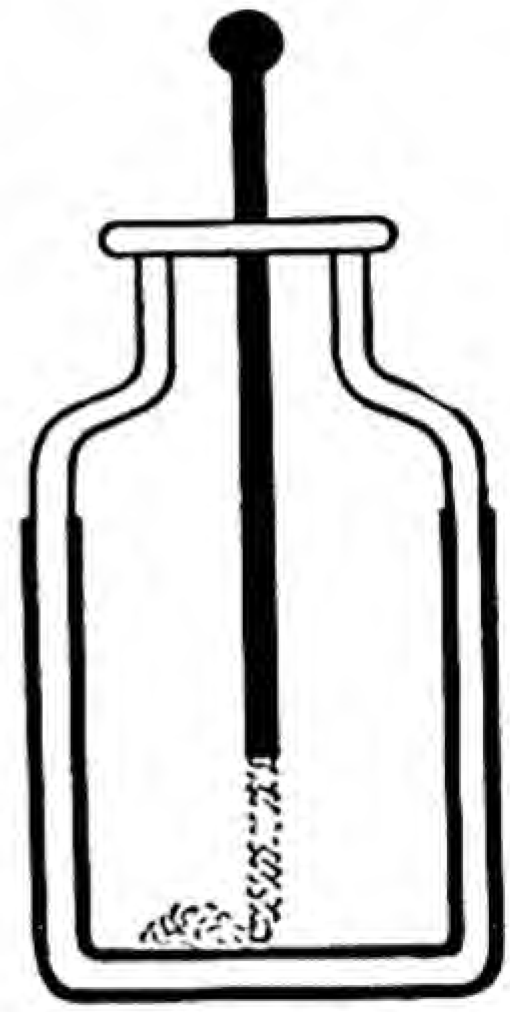


FIG. 83
Leyden Jar

Combination of Capacities

If several condensers are joined together, the capacity of the combination depends on the manner of joining. If the plates of higher potential are joined together, and also the plates of lower potential, as in Fig. 84(a), the condensers are said to be joined *in parallel*, and the combined capacity is the sum of the individual capacities. It may be gathered from the figure that the result is the same as though we had a single con-

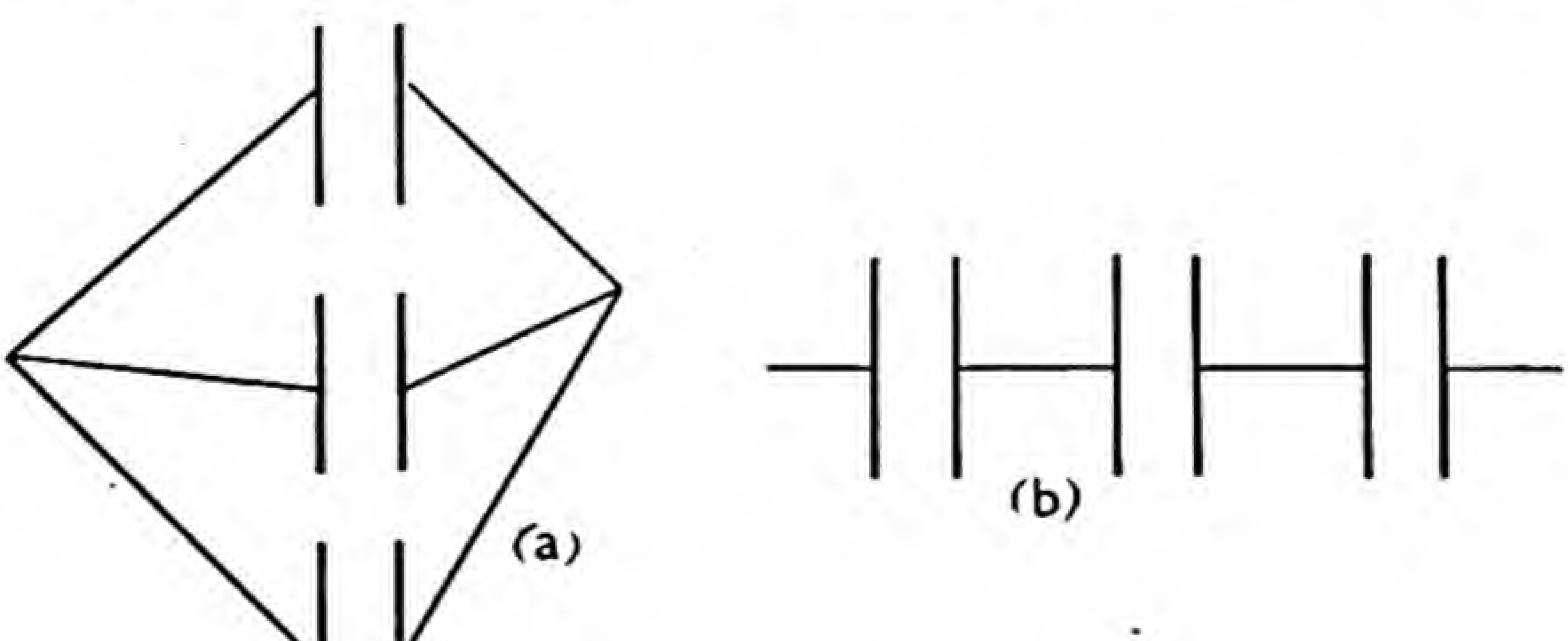


FIG. 84
Condensers joined (a) in Parallel (b) in Series

denser with a plate area equal to the sum of the areas of the individual ones, in which case the relation just stated is obvious; but it is in fact true generally, however the separation of the plates may vary from one component condenser to the next. If C is the combined capacity, and C_1, C_2 . . . the individual capacities, the relation for connection in parallel is therefore

$$C = C_1 + C_2 + \dots \qquad (9{\cdot}10)$$

If, on the other hand, the plate of lower potential of one is joined to the plate of higher potential of the next, as in Fig. 84(b), the condensers are joined *in series*. In this case the relation is

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \dots \qquad (9{\cdot}11)$$

The combined capacity is therefore greater than any of the individual capacities when the joining is in parallel, and less when it is in series.

A very common form of compound condenser in which the capacity can be varied by varying the plate area is shown in Fig. 85. One set of plates can be rotated so that any desired fraction of the area can be brought opposite the other set.

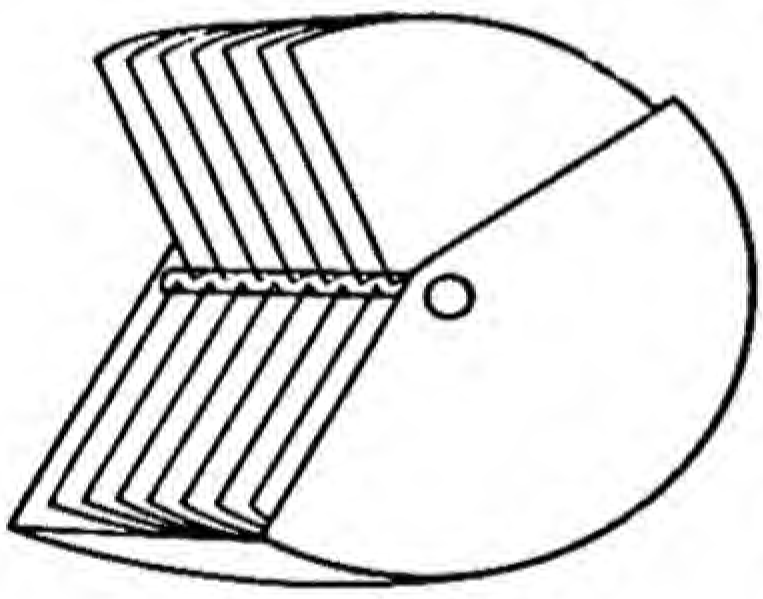


FIG. 85
Parallel Plate Condensers with Variable Capacity

Specific Inductive Capacity

It has been assumed above that the space between the conductors in a condenser contains simply air or a vacuum. If a different insulator be placed there, it is found that the capacity is changed; *i.e.* a different quantity of electricity is required to raise the potential by unity. The capacity of a condenser thus depends on the material between the plates—the *dielectric*, as it is called. It is always greater for a material dielectric, such as air, paraffin

wax, mica, etc., than for a vacuum, but the increase with gases is so small that for most practical purposes we may take the capacity as the same with air or a vacuum as the dielectric.

The ratio of the capacity of a condenser with a given dielectric to the capacity with a vacuum between the plates is called the *specific inductive capacity*, or *dielectric constant*, of the dielectric. Its value ranges for most solids between 1 and 10 (ice is an outstanding exception, having a dielectric constant of 94), and for liquids it is considerably higher, reaching a value of 81 in the nearly extreme case of water.

Specific Inductive Capacity and Electric Force

The variation of capacity with dielectric is a fact of far-reaching importance. It is found to be connected with the refractive index (pp. 61–62) of the dielectric, in a manner which we cannot here explain, but we must consider the effect of the variation on the measurement of electric force. Consider, for example, our spherical condenser (Fig. 82), of which the capacity with a vacuum between the spheres was

$$C = \frac{ab}{b-a} = \frac{a}{1-\dfrac{a}{b}} \quad . \quad . \quad . \quad . \quad (9{\cdot}12)$$

If a dielectric whose specific inductive capacity is K is used, the capacity is

$$C' = \frac{Ka}{1-\dfrac{a}{b}} \quad . \quad . \quad . \quad . \quad . \quad (9{\cdot}13)$$

Now let the radius b be increased until at last it becomes infinite. We then have our original spherical conductor, and the capacities become, as we see from (9·12) and (9·13),

$$\left.\begin{aligned} C &= a \\ C' &= Ka \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (9{\cdot}14)$$

The difference is that in the former case the sphere is in a vacuum, and in the latter it is in a space filled with the

dielectric K. It follows that, in the latter case, K times as much electricity is required to raise the potential of the sphere by unity. This can be so only if the force resisting the approach of the unit charge is reduced ; instead of being $\frac{Q}{r^2}$ at distance r it must be $\frac{Q}{Kr^2}$.

Definition of Unit Charge

Our definition of unit charge (p. 135) can therefore be generalised. We can now say that a unit charge is that charge which, when placed at a distance of 1 cm. from an equal and similar charge, repels it with a force of $\frac{1}{K}$ dynes, where K is the dielectric constant of the material between the charges. All measurements depending on the law of force must, of course, undergo a similar generalisation. For instance, the intensity of the field in the neighbourhood of charges $Q_1, Q_2, \ldots$, is the resultant of forces $\frac{Q_1}{Kr_1^2}, \frac{Q_2}{Kr_2^2} \ldots$

EXERCISES

1. Explain the meaning of electrical potential, and find the potential at one corner of an equilateral triangle, of 10 cm. side, when charges of $+ 20$ and $- 5$ units are placed at the other corners.

2. Draw the lines of force and the lines of equipotential in the field of two point charges of $+ 50$ and $- 50$ units respectively, placed 10 cm. apart.

3. Prove that the force in any direction at a point in an electric field is equal to the rate of fall of potential in that direction. Hence deduce that the direction of the resultant intensity of the field is that in which the potential changes most rapidly.

4. What is meant by the capacity of a conductor? Prove that the capacity of a spherical conductor is equal to its radius.

5. Find an expression for the capacity of a condenser consisting of two concentric spheres, the space between which is evacuated. If the space is filled with paraffin wax, how would you expect the potential difference between the spheres to be affected?

CHAPTER X

THE ELECTRIC CURRENT

WHEN a conductor joins two points of different potential, electricity flows from one to the other along the conductor until the potentials are equal. This process is very rapid, and with good conductors is completed in a fraction of a second. While it lasts, an *electric current* is said to flow from one point to the other. By convention, the direction of the current is said to be that from the higher to the lower potential, *i.e.* the direction in which positive charges would travel, but actually, owing to their much greater mobility, it is the negative electrons which move, and it is their motion which constitutes the current. It is unfortunate that, before the existence of electrons was thought of, the conventional direction of the current should have been wrongly chosen, but it is now too late to alter the convention. The student must bear in mind that when a current is said to flow from **A** to **B**, what actually happens is that electrons flow from **B** to **A**.

Contact Potential Differences

In order to maintain a current for a considerable time, the potentials of the two points must be kept different from one another in spite of the flow of electricity. The simplest way of creating a potential difference is simply to bring two conductors into contact with one another. If, for example, two metal wires are joined together, electrons from each wire can wander into the other. If the wires are of the same material the drift in one direction will be the same as that in the other, and there will be no change in the condition of either wire; but if they are of different materials (say, iron

and copper) the freedom of movement of the electrons will be different, and on the whole more will move from iron to copper than from copper to iron. The iron thus acquires a higher potential than the copper. This phenomenon is called a *contact potential difference.* Conductors can be arranged in a series such that when any two of them are placed in contact, that which appears earlier in the series acquires the higher potential. Furthermore, if several metals are joined end to end, the algebraic sum of the potential differences

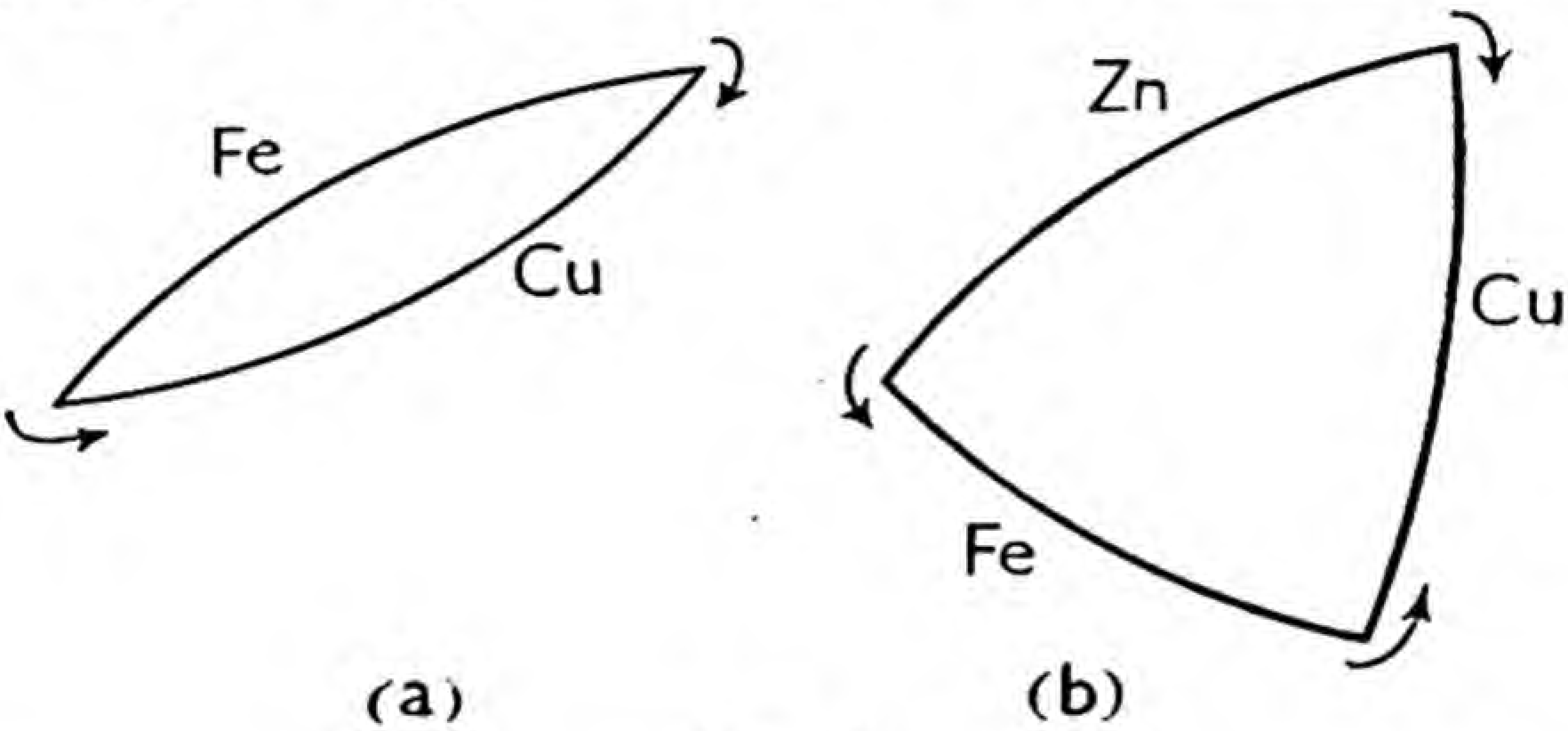


Fig. 86

Directions of Fall of Potential at Junctions of (*a*) Two, and (*b*) Three Metals

(P.D.'s) at the various points of contact is always equal to that which would be established if the first and last were directly joined. Thus, if zinc, iron, and copper are joined in that order, the sum of the P.D.'s at the Zn-Fe and Fe-Cu junctions is equal to the contact P.D. between Zn and Cu.

It is easy to see, however, that contact P.D.'s do not give us a direct means of producing an electric current. For, if two wires are joined only at one end, there is no complete conducting circuit along which electrons can flow, while if they are joined at both ends (see Fig. 86 (*a*)) the flow at one junction will be exactly neutralized by that at the other. Nor can we obtain a current by making the second junction through a third metal (Fig. 86 (*b*)), for, owing to the relation

stated above, the total P.D. in the whole circuit is again zero on balance.

Contact P.D. and Temperature : There are two means, however, of producing a permanent current in spite of this fact. The first is to keep the two junctions in Fig. 86 (*a*) at different temperatures. The flow of electrons depends on the temperature, since temperature is closely associated with the mobility of the particles of a substance, and it is found that, up to a point, the contact P.D. of two metals increases with temperature. Hence, if the two junctions are at different temperatures, the contact P.D.'s are different, and a current will flow through the circuit in the direction of the P.D. at the hotter junction. The converse of this effect also occurs. Thus, if by some other means a current is made to flow through a circuit consisting of two metals, heat is developed at one junction and absorbed at the other. These phenomena are given the general name of *thermo-electricity.*

Thermo-Electricity

Contact P.D.'s are not large, so that only feeble currents can be produced in this way, but the effect is turned to practical use in the measurement of temperature. If one junction of a *thermocouple*, as a circuit such as Fig. 86 (*a*) is called, is kept in melting ice, and the other is inserted in the body whose temperature is to be measured, the strength of the current which flows round the circuit gives a measure of the temperature of the body. (The " strength " of the current is the amount of electricity which flows per second ; instruments for measuring it, called *galvanometers*, will be described later.) For small temperature differences the Centigrade temperature of the hot junction may be taken as proportional to the current, but for larger ones a correcting term is necessary, depending on the metals used. It is usual to calibrate the couple first of all by finding the current obtained for a number of known temperatures and drawing a curve of the results. Any un-

known temperature within the range of calibration can then be read directly from the curve.

In order to increase the sensitivity of the instrument, several thermocouples are joined together in series (see Fig. 87, in which the continuous lines represent one metal and the broken ones the other) to form a *thermopile*. This affords a very delicate means of detecting and measuring small differences of temperature. It is specially useful for measuring the intensity of radiant heat. One set of junctions is coated with lampblack to give it the maximum absorbing power (see I, 185), and exposed to the source of radiation. It is heated thereby, and a sensitive galvanometer included in the circuit measures the current passing. The sensitivity of the instrument is increased if the junctions to be heated are placed in a vacuum.

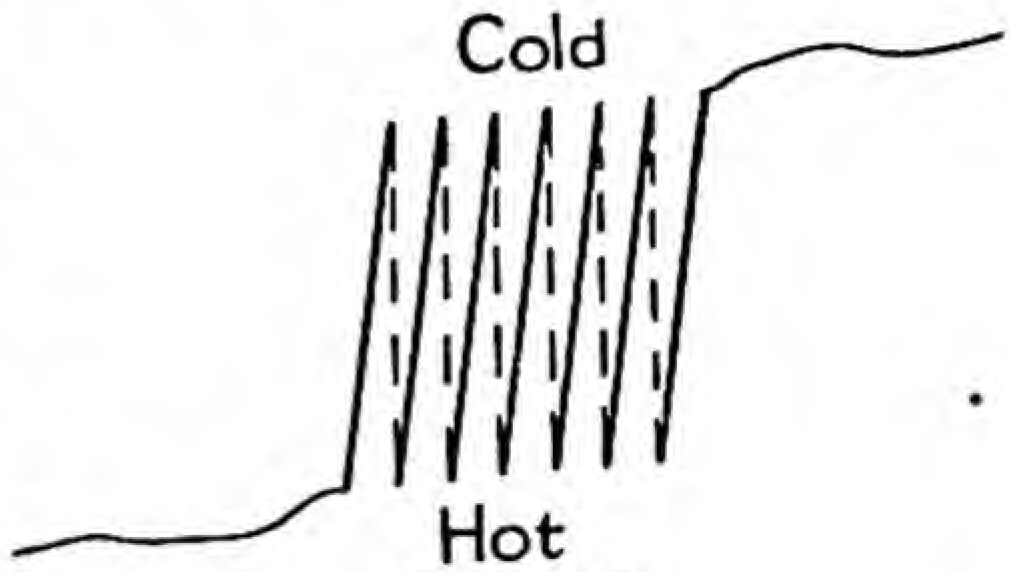


FIG. 87
Diagrammatic Representation of Thermopile
—— Metal A ---- Metal B

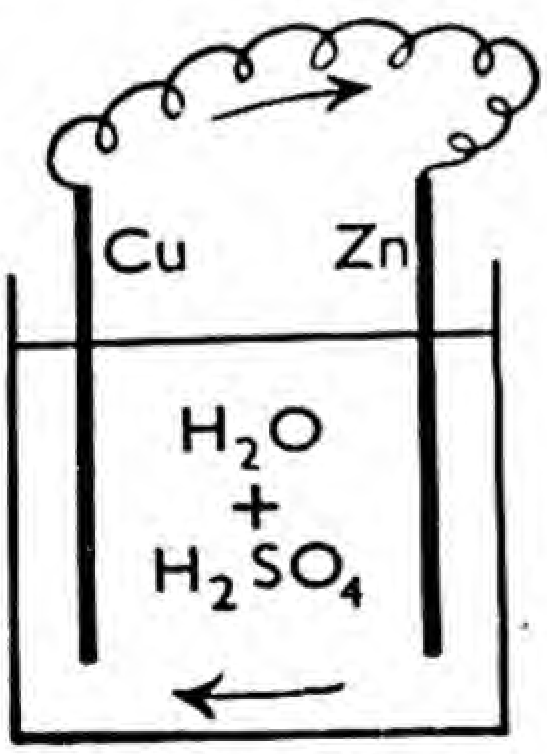


FIG. 88
The Simple Voltaic Cell

Liquid Conductors

The second means of using contact P.D.'s to give a continuously flowing current is to include a liquid conductor in the circuit. In such a case the rule that the sum of the P.D.'s at all the contacts in the completed circuit is zero at uniform temperature no longer holds. The simplest case is that of two rods of copper and zinc, respectively, placed at opposite ends of a vessel of dilute sulphuric acid (Fig. 88). In such a case the copper acquires a higher potential than the acid, and the acid a higher potential than the zinc. The direct contact P.D. between copper and zinc is comparatively negligible, so that if the copper and zinc are joined by an ex-

ternal conducting wire, a current tends to flow through it from the former to the latter. Within the acid the current flows from zinc to copper (the mechanism is explained on page 163), and a continuous flow is thus set up around the circuit.

Voltaic Cells

An arrangement of this kind, by which a continuous current is produced in a circuit containing a conducting liquid, is known as a *voltaic cell*, and the particular example just described is called a *simple cell*. The metal rods are called the *poles*, or *electrodes*, of the cell. To understand what is occurring in it we must recall our picture of the atom, and see how the relation of atoms and molecules to one another differs in solids, liquids, and gases.

Molecules in Gases

It will be remembered (see Chapter I) that the atom of any substance consists of a massive positively charged nucleus, with a number of electrons circulating round it. It is comparatively rare for the atoms of an element to exist independently of one another; usually they are associated in groups, called *molecules*. Thus, in the simplest cases (*e.g.* hydrogen) they unite in pairs forming *diatomic* molecules. The molecule of hydrogen may be represented diagrammatically as in Fig. 89, in which the large dots represent the nuclei and the small dots the electrons—which, of course, revolve round the nuclei, only instantaneous positions being shown in the figure.

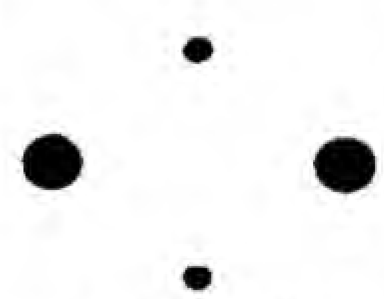


FIG. 89
Model of the Hydrogen Molecule

It is impossible to say to which nucleus a particular electron belongs; the individuality of the atoms is lost in the molecule. A jar of hydrogen gas consists of a large number of such molecules with comparatively large spaces between them —all, of course, in more or less rapid motion. A compound gas similarly consists of separate molecules; thus CO_2, for example, is made up of molecules, each having three nuclei

(one of carbon and two of oxygen) with twenty-two revolving electrons (six coming originally from the carbon atom and eight from each of the oxygen atoms), as in Fig. 90, which again is purely diagrammatic.

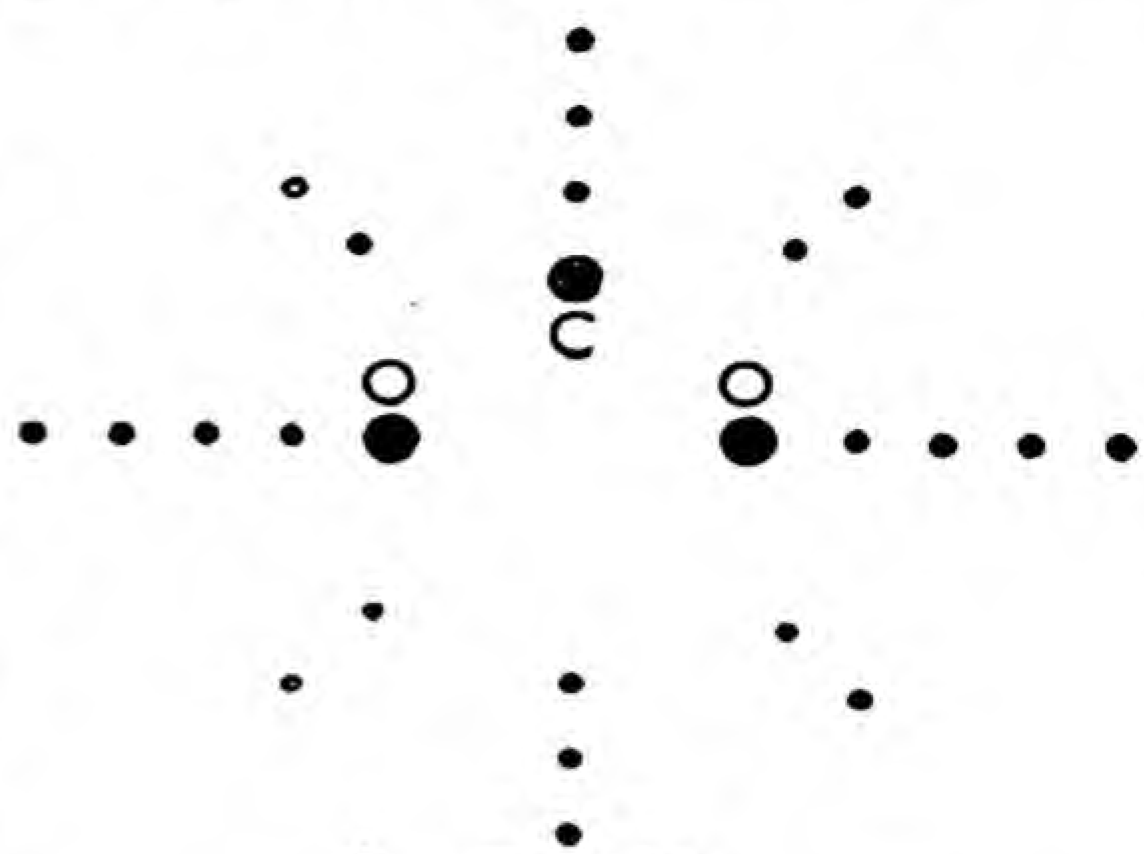


FIG. 90
Model of the Carbon Dioxide Molecule
(The distribution of the particles is purely diagrammatic)

Molecules in Solids

In a solid—especially a crystalline solid—the molecules are as close together as possible and form a whole organization in which the identity of the individual molecules is lost, just as that of the individual atoms is lost in the molecule of a gas. Thus, in sodium chloride (NaCl) the arrangement of the molecules is shown in Fig. 91, in which the closed dots represent the nuclei of sodium, and the open ones those of chlorine. Interpenetrating this

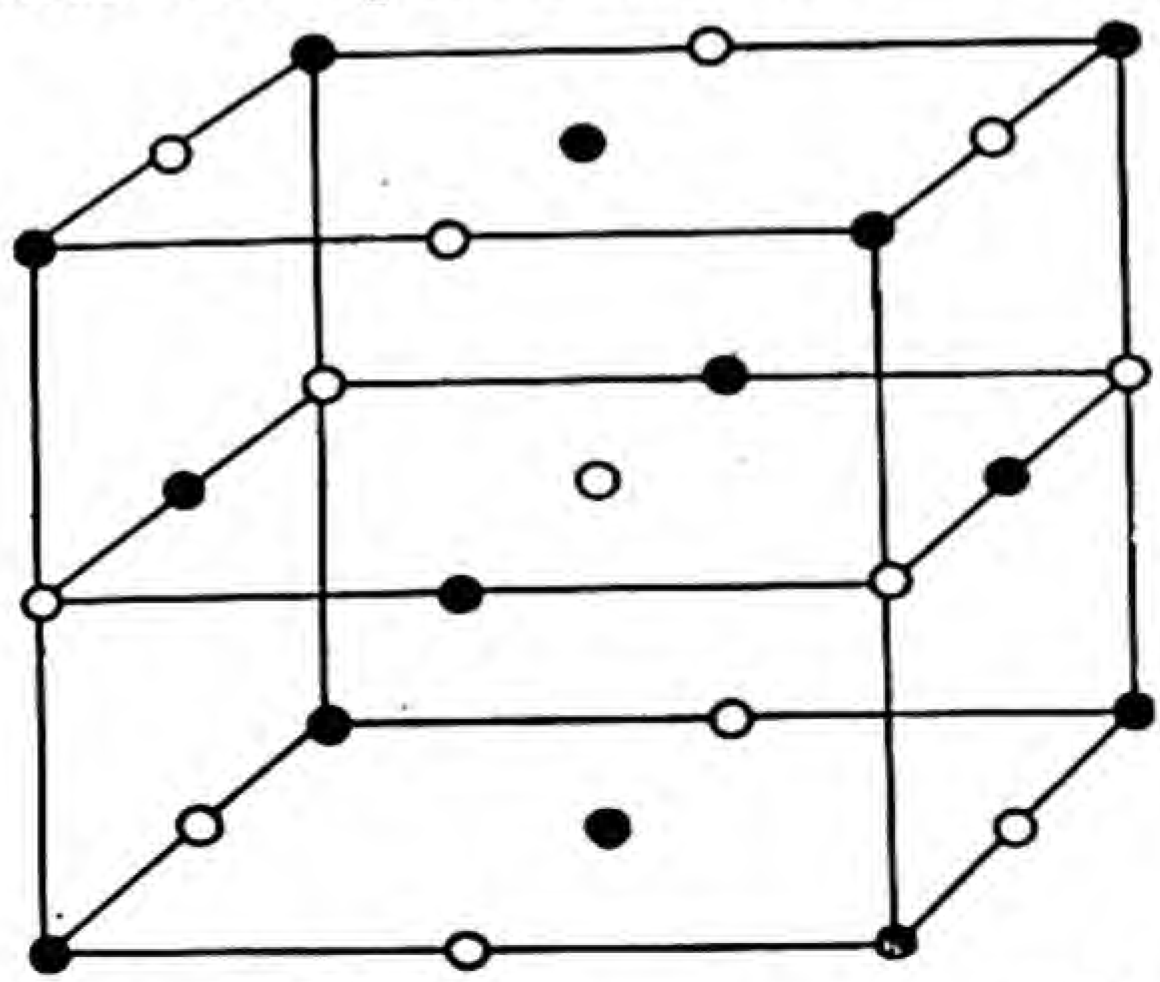


FIG. 91
Arrangement of Atomic Nuclei in Crystal of Sodium Chloride
● Sodium Nuclei
○ Chlorine Nuclei
(The straight lines are intended to make the arrangement clearer; they have no actual existence)

system there are twenty-eight electrons for each pair of nuclei; these are not shown in the figure. We may regard these electrons as holding the nuclei in position by the attrac-

tion which they exert against the mutual repulsions which would otherwise drive the nuclei apart. Here again it is impossible to say to what molecule a particular electron belongs.

Molecules in Liquids

A liquid represents a sort of intermediate state between a solid and a gas. Here the nuclei have freedom of movement, as in a gas, but there are no open spaces between the molecules so that one can be distinguished from another. The molecules appear often to form little groups which move about with respect to one another. There is a special interest, however, in liquid solutions—particularly dilute ones—because here the dissolved substance, even though originally a solid, spreads out to fill the whole volume of the solvent. The crystalline organization of Fig. 91 must break up into parts, and the electrons must choose to what nuclei they will attach themselves.

It appears that soluble substances fall into two classes, according to their behaviour when dissolved. Those in one class (called *non-electrolytes*, because they are non-conductors of electricity) break up into parts, in each of which the number of electrons is just sufficient to balance the total charge on the nuclei, and each little aggregation (it may be a single molecule or a group of molecules) is therefore electrically neutral. Sugar is an example of such a substance. Solutions of the other class (*electrolytes*, *i.e.* conductors of electricity) break up into *ions*, as they are called, each of which is a part of a single molecule and has either a positive or a negative resultant charge. Thus, when sodium chloride forms a weak solution in water it breaks up into sodium and chlorine ions. Each of the former is a sodium atom minus one electron, and each of the latter is a chlorine atom plus one electron. This is an example of the tendency mentioned in Chapter I (pp. 17–18) for each nucleus to surround itself with complete shells of electrons; the neutral sodium atom has

only one electron in the M shell, while the neutral chlorine atom has two subdivisions of that shell complete except for one vacancy. The sodium atom thereupon passes on its single M electron to the chlorine atom, forming positively and negatively charged ions. These ions attach themselves to separate groups of water molecules, and the solution when undisturbed consists of this intermingled collection of oppositely charged groups of molecules.

Conduction of Electricity

The conduction of electricity through solids consists entirely in the movement of electrons. These pass along the solid through the system of nuclei, and others follow them, so that there is never an accumulation or defect of electrons at any one place. Conduction through a liquid, however, consists of the motion of both kinds of ions in opposite directions; the positive ions move in the conventional direction of the current, and the negative ions in the opposite direction. Their movement is not so rapid as that of electrons because of their much greater mass, and liquids on the whole are worse conductors than metals.

Mechanism of Simple Cell

We can now picture what takes place in the simple cell. On completing the circuit by joining the zinc and copper externally, several contact P.D.'s are able to take effect. Sulphuric acid solution is an electrolyte, and its molecules break up into hydrogen nuclei without electrons (positive ions) and SO_4 groups with an excess of electrons (negative ions). The P.D. between copper and acid acts in the same direction as that between acid and zinc, and causes the H ions to move through the solution towards the copper and the SO_4 ions towards the zinc. Each of the former, on reaching the electrode, recovers its lost electrons from the metal and escapes into the air as a molecule of hydrogen. The

SO_4 ions, on the other hand, when they reach the zinc, combine with it to form zinc sulphate ($ZnSO_4$), which goes into the solution, and their surplus electrons travel round the external part of the circuit, and ultimately reach the copper electrode, where they are given up to the hydrogen ions, as already stated. The Zn and the H_2SO_4, therefore, gradually disappear, and, when one of them is used up, the action of the cell ceases. The pole Zn, toward which the negative ions flow, is often called the *anode*, and the opposite pole the *cathode.* The ions which move in the solution towards the anode are called *anions*, and the others *cations*.

There are several other practicable voltaic cells, *e.g.* Daniell's cell, Grove's cell, Bunsen's cell, Leclanché's cell, etc., in which different electrolytes and electrodes are used, but the principle of action in all is the same. A group of cells is often called a *battery*.

Electromotive Force

The P.D. which exists between the electrodes of a cell before they are joined externally is called the *electromotive force* (written E.M.F.) of the cell. It is an unfortunate name, since a potential difference is not a force, but a condition measured in work units. However, the name is too firmly established to be changed. Two cells—the Clark cell and the Weston cadmium cell—have very constant E.M.F.'s, and are often used as standards of P.D. for this reason.

It is important to remember two points about the electric current. First, it can flow only in a completely closed circuit of conductors. If the circuit is broken at any point the current stops everywhere in the circuit. Secondly, when the current is flowing, no matter how complex the circuit and how many changes of material and contact P.D.'s there may be, the current is the same throughout, *i.e.* if we take any imaginary complete cross-section of the circuit, the amount of electricity crossing it per second is the same wherever the cross-section is taken. If this were not so, electrons or ions would be

accumulating somewhere in the circuit, and there would be a defect of them elsewhere, and this never occurs ; the flow is continuous. Of course, if the current divides somewhere into two or more branches which unite again farther along the circuit, it is the sum of the currents along the several branches, and not the individual branch currents, that is equal to the current before it divides.

Electrolysis

If the electrodes of a cell are joined externally not by a simple wire but through a solution of an electrolyte, the passage of electricity through the electrolyte takes place by the same mechanism as that through the cell itself, and we have a similar migration of the ions. This is used in the process of *electroplating*. Suppose, for example, the electrolyte is silver nitrate. The silver ions are the cations, and when they reach the cathode they pick up some of the electrons travelling round the circuit, and form neutral atoms of silver which are deposited on the cathode. In this way the cathode becomes coated with a layer of silver. This phenomenon is called *electrolysis*.

Faraday's Law of Electrolysis: The law of operation of electrolysis was discovered by Faraday, and is known as Faraday's Law of Electrolysis ; it is as follows. *The mass of an ion deposited by electrolysis is proportional to the quantity of electricity which passes and to the chemical equivalent of the ion.* Since the current is defined as the quantity of electricity passing in unit time, the first part of the law may also be expressed by saying that the mass deposited is proportional to the current and to the time of its passage.

The dependence on the quantity of electricity is now easy to understand, though it was not so easy in Faraday's time. Each ion carries the same charge—that of 1, 2, 3, . . . electrons, according to the number of electrons it has gained or lost—so that the number of atoms deposited, and therefore the mass deposited, is proportional to the charge carried. With regard

to the second part of the law, it is to be noted that the chemical equivalent of an atom or molecule is the atomic or molecular weight divided by the valency, so that the greater the valency the smaller the chemical equivalent for a given molecular weight. Now the valency is equal to the number of electrons which are gained or lost by the atom or molecule to form an ion; a monovalent ion has gained or lost one electron, a divalent ion two, and so on. The greater the valency, therefore, the more electricity passes with each ion, *i.e.* the fewer ions for a given quantity of electricity. Thus, small chemical equivalents go with small masses deposited, and the quantities are, in fact, proportional to one another.

Electrochemical Equivalent: The mass of an ion deposited by the passage of a unit quantity of electricity is called the *electrochemical equivalent* of the ion. Faraday's law may therefore be expressed by saying that the mass of an ion deposited is equal to the product of the quantity of electricity which passes and the electrochemical equivalent of the ion. The unit of quantity chosen for this purpose is not that already defined (the electrostatic unit—E.S.U.—as it is called), but a unit which is 3×10^9 E.S.U. This unit, the derivation of which will be explained later, is known as the *coulomb*. The electrochemical equivalent of a substance is therefore the mass deposited by the passage of one coulomb.

Faraday's law may be expressed in symbols as follows. If M is the mass deposited, Q the quantity of electricity which passes ($= Ct$, where C is the current and t the time of passage), and E the chemical equivalent of the substance,

$$\text{then} \quad M = kEQ = kECt \quad . \quad . \quad . \quad . \quad (10{\cdot}1)$$

where k is a constant. If Q is 1 coulomb, then $M = kE$ and kE is the electrochemical equivalent. We can determine k by observing the mass deposited of any ion whose chemical equivalent is known, and we can then determine the electrochemical equivalent of any other such ion immediately. For

example, it is found that one coulomb deposits 0·001118 gm. of silver, whose chemical equivalent is 107·94.

$$\text{Hence} \quad k = \frac{0{\cdot}001118}{107{\cdot}94} = \frac{1}{96550} \quad . \quad . \quad . \quad . \quad (10{\cdot}2)$$

It may thus be seen that 96550 coulombs are required to deposit one *gram equivalent* (*i.e.* a number of grams equal to the chemical equivalent) of silver, and since k is independent of the element concerned, it is clear that this quantity will deposit one gram equivalent of any other ion also.

Measurement of Current

Faraday's law provides us with one method of measuring the strength of a current. Pass the current through the solution of an electrolyte for a known time (t seconds), weighing the cathode before and after the passage. The increase in weight is the weight of metal deposited. If this is n gram equivalents, then the amount of electricity which has passed is $96550n$ coulombs, and the current is therefore $96550\frac{n}{t}$ coulombs per second. A current of one coulomb per second is called an *ampere*, and a measuring instrument of this kind is called a *voltameter*. It has the disadvantage of being very slow in action, but gives very exact results, and is therefore useful for standardizing other instruments.

Electrical Resistance

Ohm's Law

Consider a wire between the ends of which a P.D. is maintained continuously by any of the means above described. We know that a current will then flow from the point of higher to that of lower potential, and we may inquire how the strength of the current changes as the amount of the P.D. is gradually increased. This question was answered by Ohm for solid conductors. *Ohm's law* states that the ratio of the

P.D. to the current is a constant quantity; it is called the *resistance* of the conductor joining the points. Thus, if C, E, and R are respectively the current, the P.D., and the resistance, Ohm's law may be expressed as

$$\frac{E}{C} = R \quad . \quad . \quad . \quad . \quad . \quad . \quad (10{\cdot}3)$$

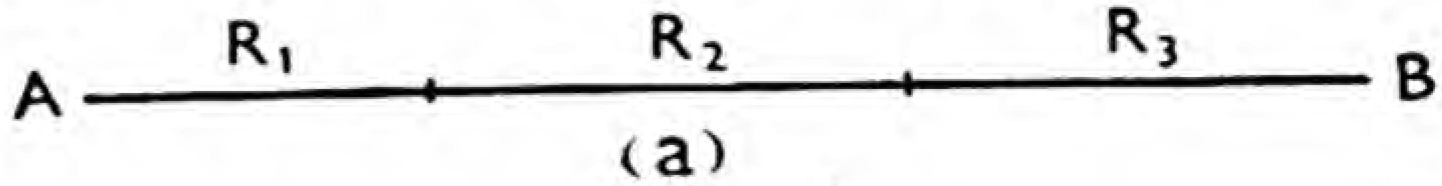
where R is constant for a given solid conductor. The units in which the quantities are measured will be explained later (pp. 209–12), but it may be said here that if C is in amperes, the most commonly used units of E and R which satisfy (10·3) are called the *volt* and *ohm* respectively.

Ohm's law is not difficult to understand in terms of the structure of solids. The greater the P.D., the greater the tendency of the electrons to move along the wire, and therefore the faster they move. The quantity of electricity passing any point of the wire per second (*i.e.* the current) is therefore increased. That it should be actually proportional to the P.D., however, cannot be rigorously proved by such simple considerations. The same reasoning applies to liquids, and provided we stop short of the electrodes, where abrupt changes of potential occur, Ohm's law holds good in electrolytes as well as in solids. In gases, however (see Chapter XIV), the conditions are different, and here Ohm's law does not apply.

Resistances in Series and Parallel

If a number of conductors are joined end to end (*i.e. in series*) their joint resistance is the sum of the individual resistances. If, on the other hand, they all join the same two points (*i.e.* are arranged *in parallel*), then the *reciprocal* of the combined resistance is equal to the sum of the reciprocals of the individual resistances. The two cases are illustrated in Fig. 92. In this respect the flow of electricity along the conductors is like the flow of liquids along pipes. If the length of the course is increased, the amount travelling round the circuit

in unit time is decreased, but if the various conductors or pipes are placed side by side, as in Fig. 92 (*b*), then there are additional channels through which the electricity or



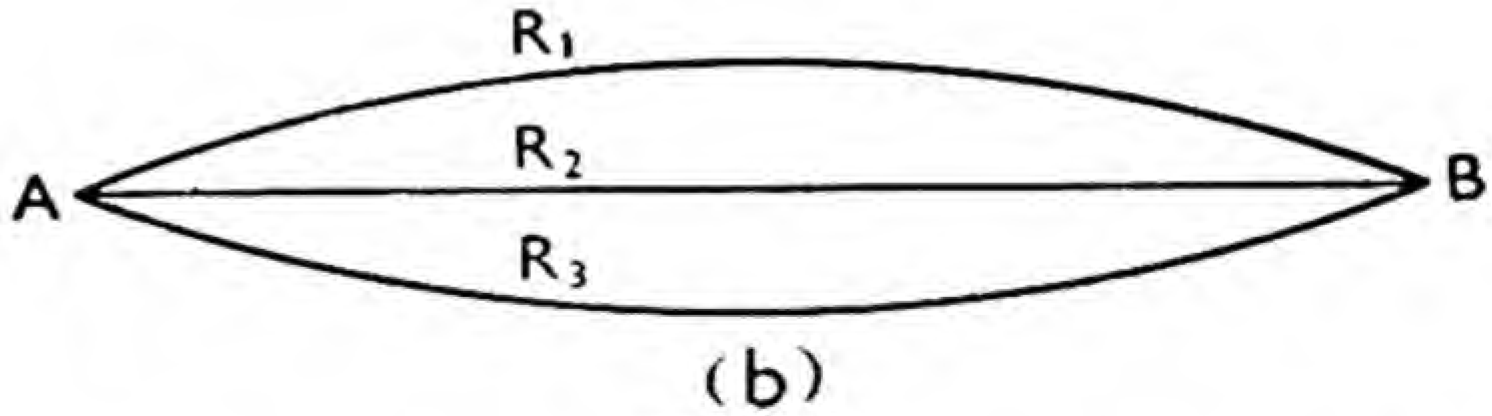


FIG. 92

Resistances joined (*a*) in series (*b*) in parallel

liquid can flow. If R_1, R_2, . . . represent the individual resistances, and R the combined resistance, then the rules just stated may be expressed as follows :

$$\left.\begin{array}{ll}\text{In series :} & R = R_1 + R_2 + \dots \\ \text{In parallel :} & \dfrac{1}{R} = \dfrac{1}{R_1} + \dfrac{1}{R_2} + \dots\end{array}\right\} \quad . \quad . \quad (10\cdot4)$$

Specific Resistance

The resistance of a single conductor depends on its dimensions, the material of which it is composed, and its temperature. It depends on the dimensions in the following way. For a given material at a given temperature, the resistance R is directly proportional to the length l and inversely proportional to the cross-sectional area a. Thus we may write

$$R = k\frac{l}{a} \quad . \quad . \quad . \quad . \quad . \quad (10\cdot5)$$

where k is a constant. It is known as the *specific resistance* of the material. This relation is quite consistent with (10·4), for we may regard a single conductor as made up of parts in two ways. First, we may divide its length AB into parts

AC and CB (Fig. 93), and regard the wire as the sum of the two conductors AC and CB in series. The total resistance is then the sum of the separate resistances, and its length is, of course, the sum of the separate lengths. Secondly, we may regard the wire as two wires in lateral contact (and therefore

FIG. 93
Resistance AB is equivalent to Resistances AC and CB in series

in parallel) of cross-sections a_1 and a_2 respectively. If R_1 and R_2 are the resistances of these two parts, then the total resistance R is given by $\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$; and if a is the total cross-section, then, putting the resistance as inversely proportional to the cross-section, we have $a = a_1 + a_2$, which is correct.

The specific resistance k varies very greatly from one material to another. Obviously, the better the conductor the smaller is k, and its reciprocal, $\frac{1}{k}$, is often called the *specific conductivity* of the material. For a perfect non-conductor $\frac{1}{k}$ is zero, so that k is infinite. There is no substance for which this is absolutely true, though for the best insulators, such as sulphur or sealing-wax, k is exceedingly high. For metals, on the other hand, it is low, and a thick bar of copper, for example, has a negligible resistance, even though of considerable length.

Resistance and Temperature

The effect of temperature on resistance is twofold. First, the dimensions of the conductor are altered by thermal expansion; and secondly, the specific resistance is changed. The first effect is comparatively small, and the main change is due to the change of k with temperature. For pure metals

the resistance increases with temperature, and this is used as the basis of a most useful thermometer, the platinum resistance thermometer. It is found that when temperature is measured on the air scale, the resistance of a metal is given by an equation of the form

$$R_T = R_0 (1 + \sigma T) \quad . \quad . \quad . \quad . \quad (10{\cdot}6)$$

where R_T and R_0 are the resistances at $T°$ and 0° C. respectively, and σ is a quantity (known as the *temperature coefficient of resistance*) which is nearly constant for moderate values of T. We thus have

$$T = \frac{R_T - R_0}{\sigma R_0} \quad . \quad . \quad . \quad . \quad . \quad (10{\cdot}7)$$

so that if σ and R_0 are determined once for all, the temperature of a body may be found immediately by measuring R_T when the thermometer is immersed in the body. Platinum is chosen because it has a very high melting-point and a large and approximately constant value of σ.

The change of resistance with temperature has disadvantages as well as advantages. A standard resistance, for example, has its standard value only at a particular temperature. To minimize the errors arising from casual changes of temperature, it is usual to make standards of alloys found by experiment to have very low temperature coefficients—*e.g.* " manganin," an alloy of copper, nickel, and manganese.

Applications of Ohm's Law

Ohm's law is so important that it is well to take a general example of its application. Suppose we have a circuit containing a number of abrupt changes of potential through contacts, cells, etc., but through which, on the whole, a current is flowing in a particular direction. In Fig. 94, wires of different materials are represented by different kinds of lines, while E_1, E_2, and E_3 represent cells of which the longer strokes indicate the positive poles. Suppose a current C flows in the direction of the arrow. (It should be noted that the

contact P.D. between an electrode and an electrolyte is always much greater than that between two metals, and the latter may usually be neglected in comparison, but we take it into account here in order to illustrate the principle.)

The variation of potential round the circuit is shown in Fig. 95. The potential at A is, of course, the same at the beginning and end of the diagram, and since the current is in the same direction all round the circuit, the slope of the graph between any two points whose abscissae are different

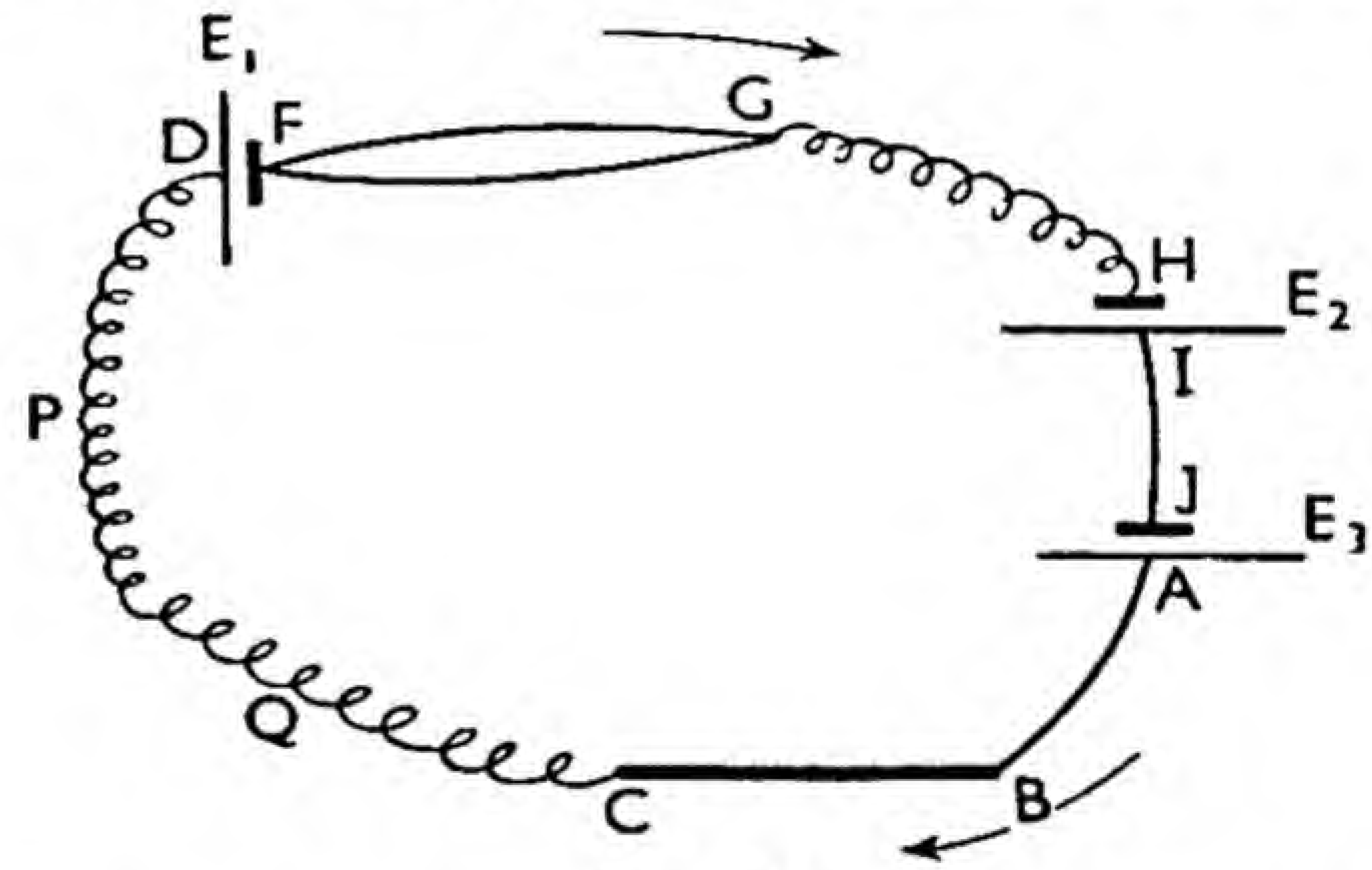


FIG. 94
Complex Circuit carrying Current in Direction of Arrows

is always in the same direction. At each junction there is an abrupt change of potential.

Let us consider first of all the application of Ohm's law to the whole circuit. To get the value of E in (10·3) we must add algebraically all the P.D.'s (the result is called the resultant E.M.F. in the circuit), and to get the value of R we must add together all the separate resistances (of these the resistance—R_0 say—between F and G is given by $\frac{1}{R_0} = \frac{1}{R_1} + \frac{1}{R_2}$, where R_1 and R_2 are the resistances of the two wires joining these points). Then the current flowing through the circuit is given by $C = \frac{E}{R}$.

Consider next a portion of the circuit—say, the portion PQ—in which there is no abrupt change of potential. Then we can determine the P.D. ($= V$) between P and Q if we know the resistance R' of PQ. For the current C has already been found, and we have therefore $V = CR'$. In this way we can determine how the potential varies along the circuit and so construct Fig. 95. We cannot, however, apply this

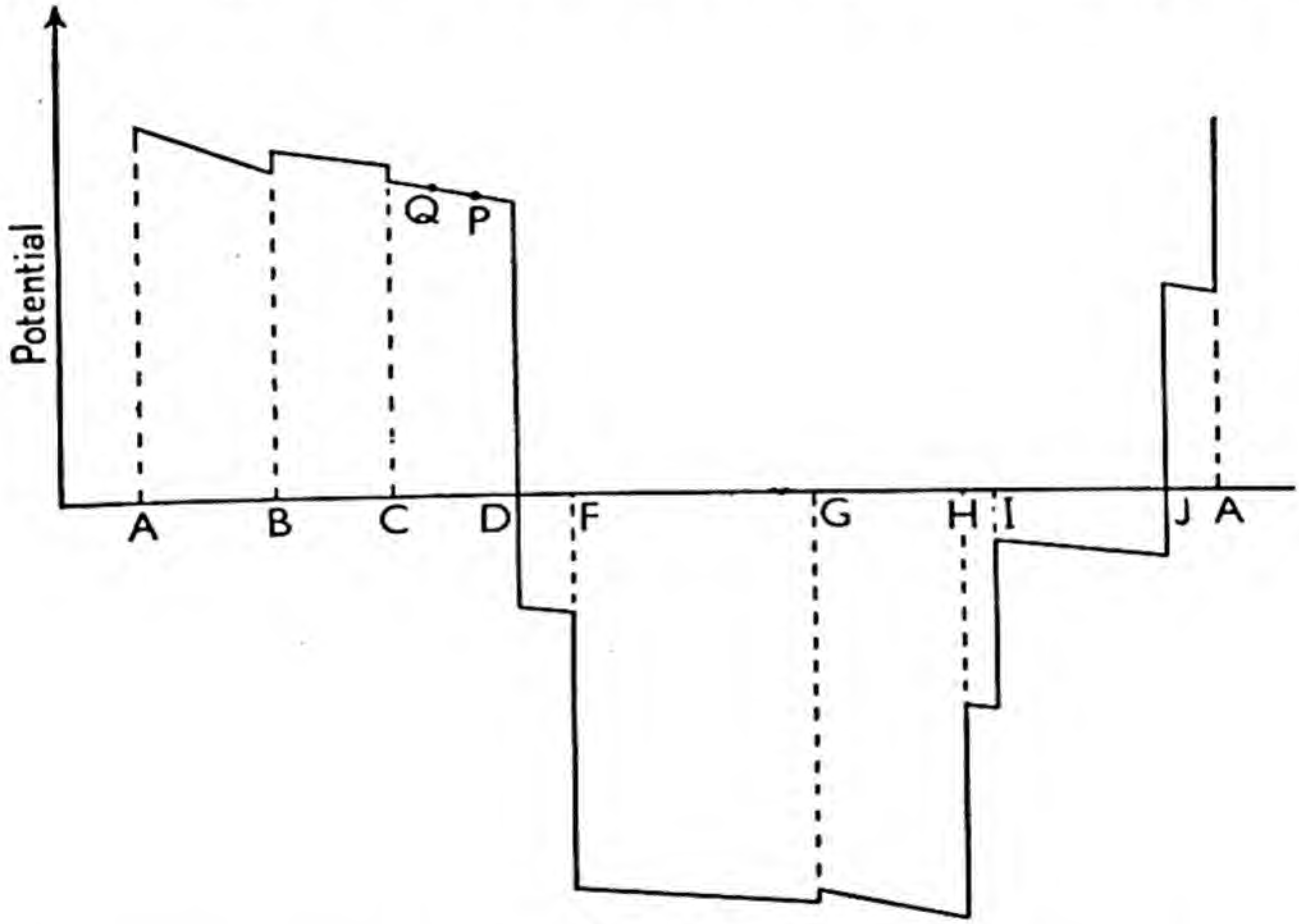


FIG. 95

Variation of Potential along Circuit shown in Fig. 94

method to two such points as P and G, between which there is an abrupt change of potential; it is valid only in the portions of the circuit in which the current varies continuously. The region within an electrolyte satisfies this condition as well as each of the single wires of the circuit, so long as the junctions are excluded.

Between F and G the current C flows partly along one path and partly along another. We can find the amount in

each path from Ohm's law, for since R_1 and R_2 are the two resistances, the currents are $\frac{V'}{R_1}$ and $\frac{V'}{R_2}$ where V' is the P.D. between F and G. The sum of these two currents is, of course, C, as we may easily verify; for it is $V'\left(\frac{1}{R_1}+\frac{1}{R_2}\right)$ *i.e.* $\frac{V'}{R_0}$ as we have already seen, and since R_0 is the net resistance between F and G, this quantity must be C.

Circuit containing Single Cell: Now let us leave this complex circuit and consider a single cell, whose terminals we can join, if we wish, by a wire of resistance R. Before we so join them, the P.D. between the electrodes is the E.M.F. of the cell; let us call this E, and let the internal resistance of the cell be r. No current, of course, now flows, but when the electrodes are joined, the circuit is completed and we have a current C given, as before, by

$$E = C\,(R + r) \quad . \quad . \quad . \quad . \quad . \quad (10{\cdot}8)$$

The P.D. between the terminals A and B outside the cell (Fig. 96) is, by Ohm's law, equal to CR, and this is less than E by the amount Cr. The effect of joining the terminals is therefore to decrease the P.D. between them. This, of course, would be expected, because electricity is now able to flow from one electrode to the other, thus tending to equalize the potentials.

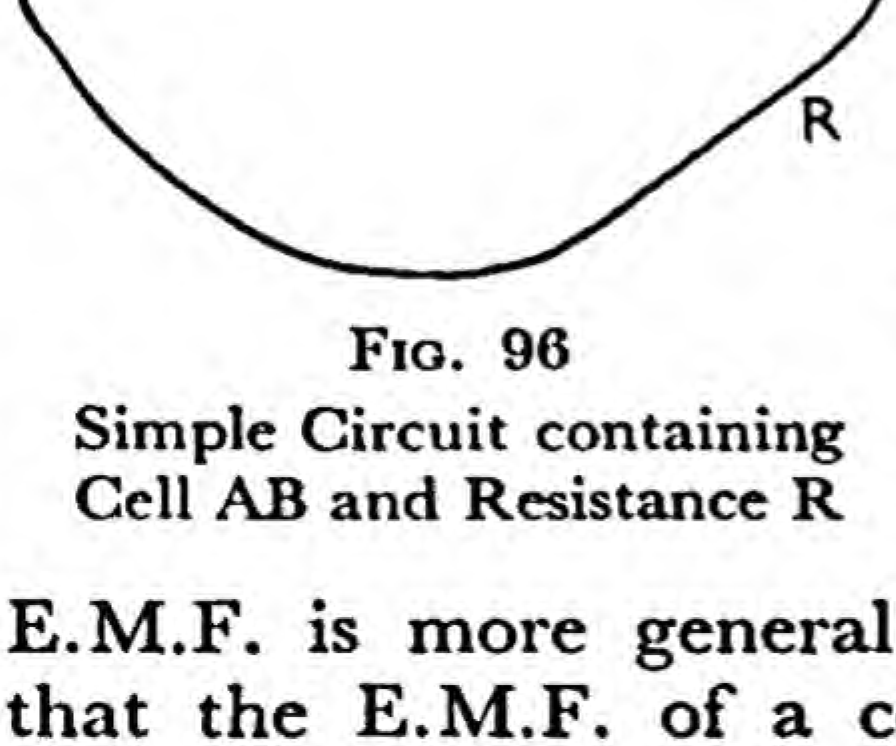


Fig. 96
Simple Circuit containing Cell AB and Resistance R

E.M.F. and P.D.

This example helps us to understand the differences between E.M.F. and P.D. The idea of E.M.F. is more general than that of P.D. When we say that the E.M.F. of a cell is two units, we mean that the P.D. between the terminals of the cell when they are not

joined by a conductor is two units. If they are joined the P.D. is reduced, but the E.M.F. remains the same. Furthermore, there are circumstances (see p. 220) in which a current can be made to flow in a perfectly continuous circuit without any abrupt change of potential—a uniform circular metal ring, for example. In such cases we say that an E.M.F. is created in the ring, but we cannot without ambiguity apply the idea of P.D. For, if we do we must say that, starting at any arbitrary point and going round the circuit in the direction of the current, the potential is falling all the time. When, therefore, we come again to our starting-point, we find that it has both its original potential and a lower one. It is best to regard E.M.F. as a tendency for electricity to flow, which is sometimes, but not always, identical with the P.D. between two points in the circuit, and is always measurable in terms of the same units as P.D.

Networks of Conductors

Ohm's law enables us to find out how a current is distributed through a network of conductors of any degree of complexity, such as that shown in Fig. 97, for example. Two laws, known

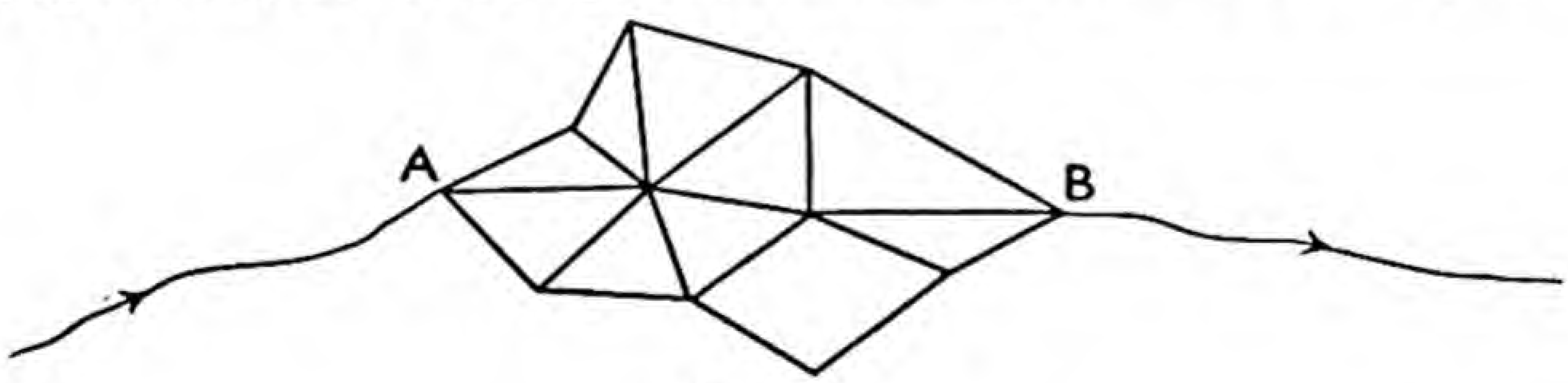


FIG. 97
Network of Conductors

as *Kirchhoff's laws*, express the way in which it is applied. The first law says that the algebraic sum of the currents meeting at any junction in the network is zero, currents flowing towards the junction being counted as positive and those flowing away from it as negative. This must be so, otherwise electricity would be continuously accumulated at or withdrawn from such points, which is contrary to experi-

ence. The second law says that in any closed mesh of the network, the total E.M.F. is equal to the sum of the products CR for each wire in the mesh.

Wheatstone's Bridge: We shall consider one particular network of great importance, known as *Wheatstone's bridge,* which is so simple that we need not apply Kirchhoff's laws in their general form, but can apply Ohm's law directly. This network is illustrated in Fig. 98. The current from the cell E divides at A into two parts, along ACB and ADB respectively, and the resistances of the various parts of the circuit are as marked in the figure. Then, if C is the total current in the circuit, this divides at A into C_1 along ACB and C_2 along ADB.

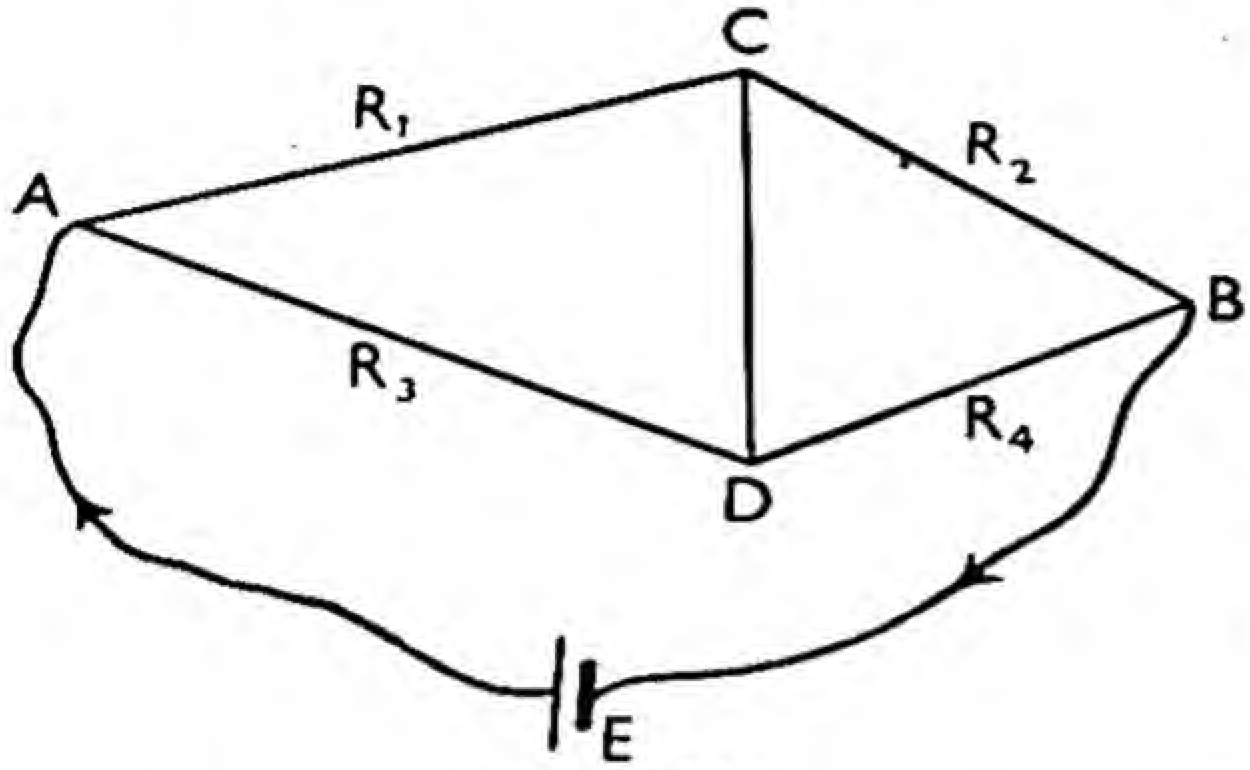


FIG. 98
Divided Circuit illustrating Wheatstone's Bridge

Let the potentials at A, B, C, and D be V_A, V_B, V_C, and V_D respectively. Then

$$\left.\begin{aligned} V_A - V_C &= C_1\, R_1 \\ \text{and } V_A - V_D &= C_2\, R_3 \end{aligned}\right\} \quad \ldots\ldots \quad (10{\cdot}9)$$

Now suppose that C and D have the same potential; then, from (10·9),

$$C_1\, R_1 = C_2\, R_3 \quad \ldots\ldots \quad (10{\cdot}10)$$

$$\text{and therefore } \frac{C_2}{C_1} = \frac{R_1}{R_3} \quad \ldots\ldots \quad (10{\cdot}11)$$

But in that case the P.D. between C and B is the same as that between D and B, so that

$$C_1 R_2 = C_2 R_4 \quad \ldots\ldots \quad (10{\cdot}12)$$

$$\text{i.e. } \frac{C_2}{C_1} = \frac{R_2}{R_4} \quad \ldots\ldots \quad (10{\cdot}13)$$

Hence, from (10·11)

$$\left.\begin{aligned} \frac{R_1}{R_3} &= \frac{R_2}{R_4} \\ \text{or} \quad \frac{R_1}{R_2} &= \frac{R_3}{R_4} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (10{\cdot}14)$$

Measurement of Resistance

The value of this result is that it affords a simple way of finding the ratio of two resistances. Suppose we have a framework consisting of thick copper bars, PLM, NS, TVQ (Fig. 99), whose resistance is negligible, the ends P and Q being connected by a straight uniform conducting wire.

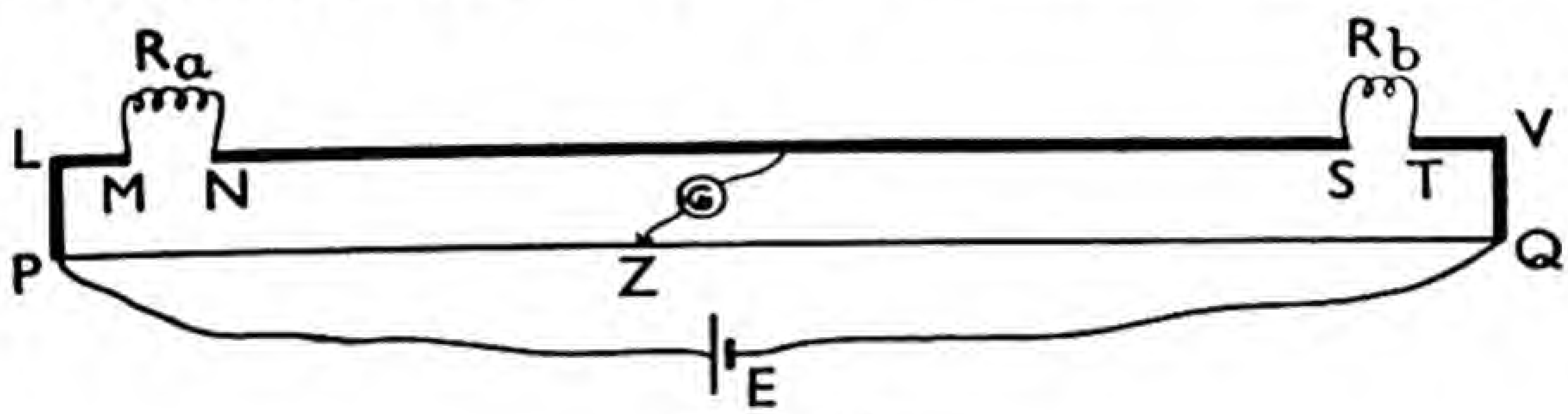


Fig. 99

The Metre Bridge

Let the resistances R_a and R_b to be compared be placed one in each of the gaps MN and ST, so that a complete circuit is formed, and let a current from a cell E pass from P to Q partly along the wire and partly through the framework and resistances. Connect one terminal of a galvanometer, or any instrument which indicates the passage of a current, to the bar NS, and move a conducting lead from the other terminal along the wire PQ until a point Z is found at which the galvanometer shows that no current is passing. Then Z must be at the same potential as NS. We then have the conditions of Fig. 98, in which $R_1 = R_a$, $R_2 = R_b$, and R_3 and R_4 are respectively the resistances of PZ and ZQ. Hence from (10·14)

$$\frac{R_a}{R_b} = \frac{\text{Resistance of PZ}}{\text{Resistance of ZQ}} \quad . \quad . \quad . \quad (10{\cdot}15)$$

But since PQ is a uniform wire of a single material, the

resistances of PZ and ZQ are proportional to their lengths, from (10·5). Hence

$$\frac{R_a}{R_b} = \frac{\mathrm{PZ}}{\mathrm{ZQ}} \quad . \quad . \quad . \quad . \quad . \quad (10{\cdot}16)$$

In order to compare two resistances, therefore, we have simply to place them in the gaps MN and ST, and move the free end of the galvanometer lead along PQ until a point Z is found such that no current flows through the galvanometer. An arrangement of this kind is often called a *metre bridge*, since PQ is usually a metre long and has a metre scale placed underneath it to facilitate measurement of the lengths concerned.

The ratio of the two resistances having thus been found, the value of either is determined if that of the other is known. Hence, by making R_a a standard resistance we can measure an unknown resistance R_b. The greatest accuracy is obtained when the standard is chosen approximately equal to the unknown resistance, so that Z is near the centre of PQ. The Wheatstone bridge is used in conjunction with the platinum resistance thermometer in the measurement of temperature. The thermometer is placed in one of the gaps, and the standard resistance in the other is chosen so that a balance is obtained near the centre of the wire.

Heating Effect of Current

When a current passes through a conductor the conductor is heated. The motion of the electrons adds to the kinetic energy of the particles composing the conductor and the temperature therefore rises. Conductivity for heat is, in fact, closely related to conductivity for electricity.

This is an example of the law of conservation of energy. Consider two points, A and B, in a wire through which a current flows, and let their potentials be V_A and V_B respectively. Then, from the definition of potential difference, one unit of electricity in passing from A to B does work equal

to $V_A - V_B$ units. If a current C flows through the wire, then C units of electricity pass per second, so that $C(V_A - V_B)$ units of work are done each second. Now, work is a form of energy, and it cannot be lost. What happens is that it is transformed into heat in the wire.

Putting $V_A - V_B = E$, we see that if the current passes for t seconds, the heat generated is

$$H = \frac{ECt}{J} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10 \cdot 17)$$

where J is the mechanical equivalent of heat (see I, 131). By Ohm's law we may write this equation in either of the alternative forms

$$\left.\begin{aligned} H &= \frac{E^2 t}{RJ} \\ H &= \frac{C^2 Rt}{J} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (10 \cdot 18)$$

where R is the resistance between A and B.

We shall deal later with the units in which these quantities are usually measured. For the moment it will be sufficient to say that a current of one ampere, passing between points differing in potential by one volt, will generate 10^7 ergs (1 joule) of work per second, and the heat produced in the wire in each second will therefore be

$$H = \frac{10^7}{J} \text{ calories} \quad . \quad . \quad . \quad . \quad . \quad (10 \cdot 19)$$

We may thus calculate the rate at which heat is produced, and arrange our apparatus to obtain the most suitable result possible in given circumstances. Suppose, for instance, that E, the P.D., is fixed, and that two wires in parallel join the points A and B. Then, from the first equation of (10·18) we see that the rate of heat production is inversely proportional to R, so that the wire with the smaller resistance is heated more. On the other hand, if the current is fixed, we see from the second equation of (10·18) that the rate of heat production is *directly* proportional to the resistance. Thus, if a circuit

contains equal lengths of a thick and a thin wire of the same material in series, more heat will be produced in the latter. Since its heat capacity also is smaller, it will rise to a much higher temperature, and may become red hot while the other wire is only slightly warmed.

The heating effect of an electric current is turned to practical use in electric lamps for lighting and electric heaters for domestic heating and cooking. It is used also to heat clothing for use at great altitudes, especially when flying. Insulated wires are placed inside thick fur coats, and a current of the proper strength is sent through them to produce the desired degree of heat.

The Electric Arc

Consider a circuit containing a number of cells so as to produce a large P.D. (100 volts or more) and two conducting rods, A and B (Fig. 100). A current passes, producing some heat in the rods (the *poles*, as they are often called). Now suppose the rods are gradually separated. The circuit is broken, but a small amount of vapour is produced from the rods which bridges the gap, and, being a conductor, enables the current to continue. Electrons pass much more readily through the vapour than through the solid, and they impinge on the end of the pole towards which they move with such energy that they quickly raise it to a white heat. This, of course, produces much more vapour, and the current is able to pass when the rods are as much as a few centimetres apart. The space between the poles is filled with an intensely bright mass of glowing vapour, and the poles themselves, especially the positive pole, also glow brightly. This is called an *electric arc*.

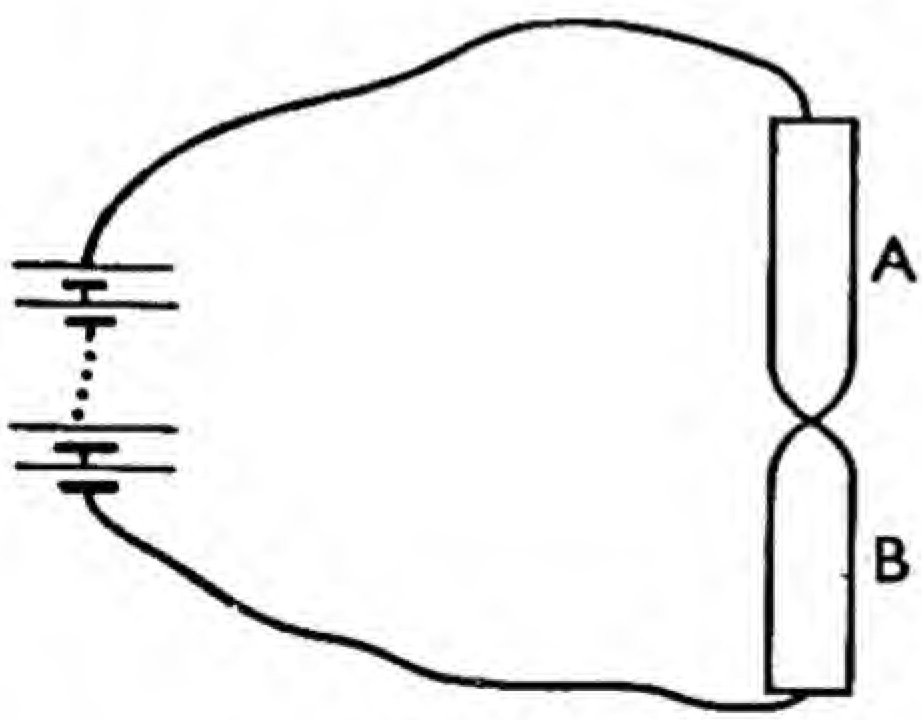


FIG. 100
Apparatus for the Electric Arc

As the arc continues to burn, it is noticed that the positive pole is gradually eaten away, a hollow called a *crater* being formed in it, while the negative pole disappears much more slowly. The material of both poles is vaporized, and so gradually disappears into the air, but the wearing away of the negative pole is partly counterbalanced by the accession of material transferred from the positive pole by the passage of the current. The atoms or molecules are ionized, and the vapour consists of some free electrons and some ionized atoms which are positively charged. The electrons move from the negative to the positive pole, and go round the circuit to keep the current going, while the ionized atoms go from the positive to the negative pole, where they pick up electrons from those carrying the current, and, as neutral atoms, add to the material of that pole.

The temperature of the arc is one of the highest producible terrestrially ; it reaches a few thousand degrees Centigrade. All known materials are vaporized at this temperature, and a substance placed on the lower pole adds its vapour to the arc, and its spectrum, or the spectra of its constituent atoms, appears when the light from the vapour is analysed in the spectroscope. This is one of the most satisfactory ways of producing the characteristic spectrum of an element which is solid or liquid at ordinary temperatures.

EXERCISES

1. Explain how an electric current can be produced by means of the contact P.D.'s between two metals. How can this be utilized in the measurement of radiant energy ?

2. Describe the structure of solids, liquids, and gases, respectively, in terms of the molecules which compose them. How are the differences between them related to the manner in which they conduct electricity ?

3. Explain the principle of electrolysis, and state Faraday's

law. If a current of 2 amperes deposits 8 gm. of silver electrolytically in an hour, what is the electrochemical equivalent of silver ?

4. State Ohm's law. If two points, maintained at a P.D. of 10 units, are joined by three wires in parallel of the same material and cross-section but having lengths in the ratios 1 : 2 : 3, find the relative values of the currents in the wires. If the wires have all the same length, but diameters in the ratios 1 : 2 : 3, what are then the relative values of the currents ?

5. Using Ohm's law, prove formulae (10·4) for the resistances of conductors in series and in parallel.

6. Explain how the variation of resistance with temperature can be applied to the measurement of temperature. If the resistances of a platinum thermometer at 0° C. and 100° C. are R_0 and R_{100} respectively, prove that the temperature at which the resistance is R_T is given by $\dfrac{R_T - R_0}{R_{100} - R_0} \times 100$.

7. The E.M.F. and internal resistance of a battery are 12 volts and 3 ohms respectively. When the terminals are joined by a wire, the P.D. between them falls to 10 volts. Find the resistance of the wire and the current through the circuit. How much heat is developed in the wire per minute ?

8. Describe Wheatstone's bridge, and explain how you would use it to determine the ratio of the specific resistances of two materials.

CHAPTER XI

MAGNETISM

Permanent Magnetism

THE electric current consists simply of electrons or ions travelling round and round a circuit, and it may well be asked why, apart from the general thirst for knowledge, we should be interested in constructing vast machines in order to make invisible particles do the same thing over and over again. Two reasons have already been given : we can use such a process to produce chemical action, as in electrolysis and electroplating, and we can use it to produce light and heat. A third, and the most important reason of all, is that we can use it to produce magnetic force. It has already been said that a circular current acts as a magnet, but before considering the magnetic effects of a current in more detail we must examine the properties of the so-called " permanent " magnets—pieces of iron and steel which attract other pieces of iron and steel without any obvious connection with electricity at all, although, as already stated, we believe the force to be associated with the motion of electrons within the magnets.

The Magnetic Field

We have seen in Chapter I that a magnet (for simplicity we shall confine our attention here to " bar " magnets, *i.e.* straight rods of iron or other magnetizable material) always has two opposite poles—a N. and a S. pole—at or near the two ends, and that it exerts a force on pieces of iron or steel in its neighbourhood. In this it resembles a pair of electrified particles, one positive and the other negative, placed at a

certain distance apart, and experience shows that the two poles always have the same strength—*i.e.* a particle of unmagnetized iron placed near one pole is attracted to that pole with the same force as it is attracted to the other pole when placed in a corresponding position with respect to it. We may therefore represent the force in the space surrounding the magnet (the *magnetic field*) by a diagram similar to Fig. 101. Here the lines of force outside the magnet are

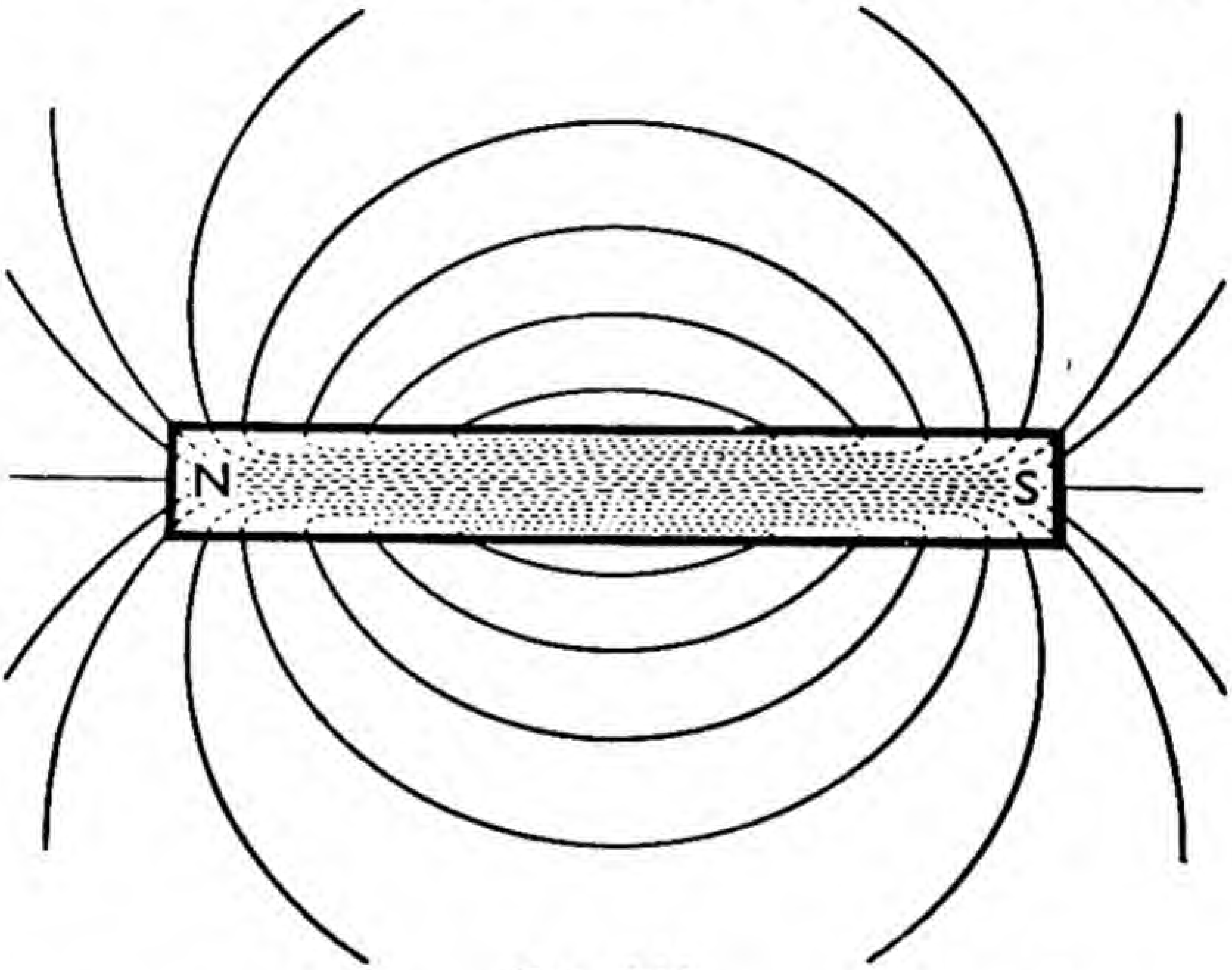


FIG. 101
Closed Lines of Force in and around Bar Magnet

practically identical with those surrounding two equal and dissimilar point charges of electricity (Fig. 76(*a*)). The conditions inside the magnet are not so easy to determine, but there is reason to believe that each line of force is continued there so as to form a closed loop in the manner shown by the dotted lines. The external attraction exerted by the magnet is, accordingly, greatest at the poles, and diminishes as we proceed inwards, becoming zero at the centre of the magnet. Although, as Fig. 101 shows, the field is quite sym-

metrical, the lines of force are conventionally regarded as issuing from the magnet at the N. pole and returning into the S. pole. This is, of course, the path which a free N. pole, if it could exist, would take in this field.

Magnetic Lines of Force

A very convenient way of displaying these lines of force is to lay the magnet on a table, cover it with a large sheet of paper, and sprinkle on the paper some iron filings. Gentle tapping then enables the filings to set themselves in obedience to the force acting on them (the object of tapping is to remove each filing momentarily from the paper, so that friction does not prevent it from setting itself along the direction of the resultant force. On falling back to the paper it retains that orientation). In this way the filings form a visible map of the lines of force indicated in Fig. 101. Each of them becomes a small magnet by induction. Its S. pole is pulled towards the N. pole, and repelled from the S. pole of the bar magnet, while its N. pole is treated in the opposite manner, with the result that it lies along the line of force.

The Inverse Square Law

In mapping the electric field we saw that we could obtain the magnitude and direction of the force at any point by placing in imagination a unit positive charge at that point and finding the resultant of the forces which the two charges would exert on it. We can do the same thing with the field surrounding a magnet, substituting a unit N. pole for a unit positive charge, for it is found by experiment that the force here also varies as the square of the distance from a pole. This rule, however, is less definite in magnetism than in electricity, for two reasons. First, the exact location of the poles is somewhat uncertain. The best we can do is to suppose that they are situated at the places where the external lines of force would meet inside the magnet if produced in the

directions in which they enter it. In that case we find that the poles are not exactly at the ends of the magnet, but a short distance inside. At best, however, the whole idea of "poles" is merely an approximation to the facts. It is more accurate to think in terms of the closed lines of force and use the conception of poles only to simplify calculations when precision is not important.

The second reason for the indefiniteness of the "inverse square" law in magnetism is that we cannot produce a unit N. pole by itself; it must always be on a piece of material having an equal S. pole at the other end, and this S. pole will change the field to be measured. For practical purposes its effect can be minimized by using a long thin magnet (*e.g.* a knitting-needle), whose S. pole is too far away to have much influence.

The Unit Magnetic Pole

We have spoken of a "unit N. pole." The unit here is defined in a manner similar to that of the unit of electric charge. A unit N. pole is that pole which, when placed at

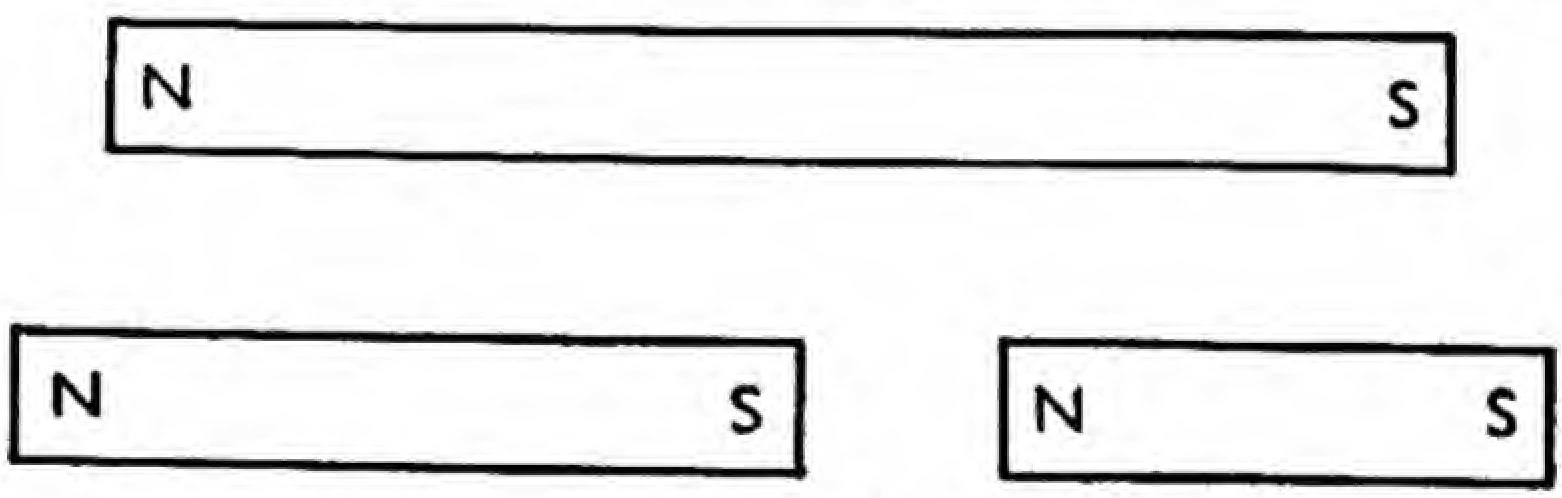


FIG. 102
Polarity developed by cutting Bar Magnet in two

a distance of one centimetre from a similar and equal pole, repels it with a force of one dyne, the S. poles of the magnets concerned being too far away to be effective. It then follows, just as with electric fields (pp. 138–39), that if we represent the strength of a field by the number of lines of force crossing unit area in it, at each unit pole 4π lines of force will meet.

To make sense of this definition, we must be able to test whether the two poles concerned are " equal " or not. We may do this by comparing their effects on the same third pole ; equal poles produce equal effects. Another method is to make use of the fact that if we cut a magnet in two, poles will be formed at the two new ends equal in strength to those of the original magnet. Thus in Fig. 102, if the upper magnet is cut into the two pieces shown below, each will be a magnet with the same pole strength as the original one, and the polarity will be as shown.

Magnetic Screening

In the method of mapping the magnetic field by iron filings, it is assumed that the filings merely reveal the direction of the lines of force without changing them in any way. This

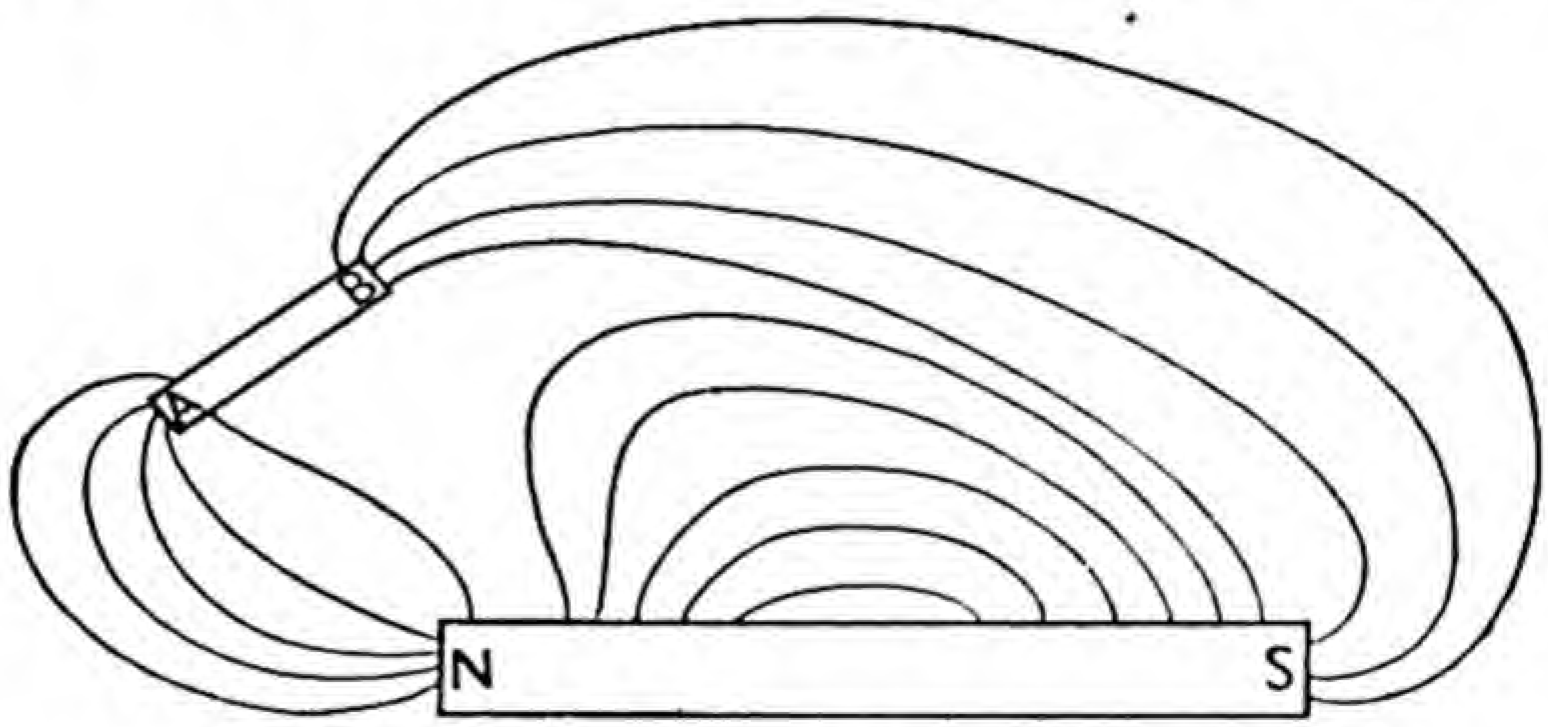


FIG. 103

Effect on Lines of Force of placing Iron Rod AB in Field of Bar Magnet NS

is not exactly true, although for tiny pieces of metal like filings, it is true enough for practical purposes. In general, however, when a piece of iron is placed in a magnetic field, the lines of force tend to crowd into it. This is illustrated in Fig. 103, where the effect of placing a piece of iron AB in the field of the magnet NS is shown. It will be seen that the space near AB is somewhat cleared of lines of force ; *i.e.* the field is made much weaker there. This effect is useful

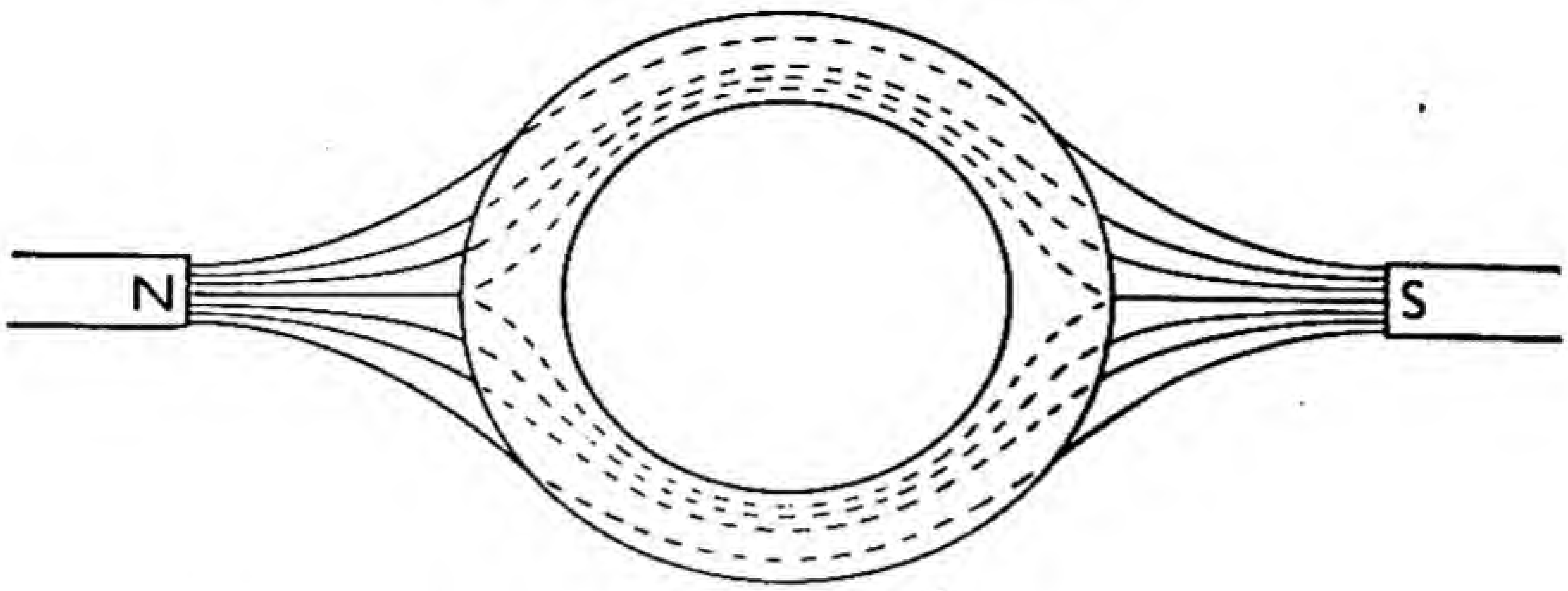


FIG. 104
Magnetic Screening of Space inside Iron Ring

when it is desired to screen a space from the influence of a magnet. A ring of soft iron, as in Fig. 104, removes the lines of force from the space inside it, and that space is therefore screened from magnetic action.

Terrestrial Magnetism

The distribution of intensity in the field surrounding a magnet is similar in almost every respect to that surrounding two equal and dissimilar electric charges. There is, however, a difference in the behaviour of bodies placed in the field, for in the electric field a free charge moves along a line of force towards one of the charges, whereas in the magnetic field a small magnet does not move as a whole, but sets itself to lie along a line of force. The tendency of the N. pole to move in one direction is checked by that of the S. pole to move in the opposite direction, and the magnet thus takes up a definite alignment with a slight strain in it arising from the opposite pulls on the two ends.

Now it is found that if a magnet be suspended so as to be able to swing freely in a place far from the neighbourhood of other magnets, it always comes to rest in a plane almost passing through the north and south terrestrial poles. The explanation of this, first given by William Gilbert of Colchester in the sixteenth century, is that the Earth itself is a magnet,

surrounded like other magnets by a magnetic field, and the swinging magnet sets itself along the lines of force of this field just as an iron filing sets itself along a line of force of the field of an artificial magnet.

Magnetic Elements: The lines of force of the Earth's field are not at most places horizontal, and a magnetic needle, free to move in any direction, will therefore point, not towards the north and south points of the horizon but somewhere in a vertical plane almost passing through those points. It does not quite pass through the north and south points, and this means that the magnetic poles of the Earth are not exactly at, though they are near, the geographical poles. The angle which the Earth's lines of force make with the horizontal at any place is called the *magnetic dip* of the place. It varies slowly with time, being nearly 70° at London at the present time. In Fig. 105, F represents a line of force of the Earth's field, and θ the angle of dip. H and V are respectively the horizontal and vertical components of the Earth's field. In most problems we are concerned with magnetic needles which are free to move only in a horizontal plane, and the effective force acting on them is then H and not F. When we speak of the Earth's field we shall in future, unless the contrary is stated, mean only the horizontal component.

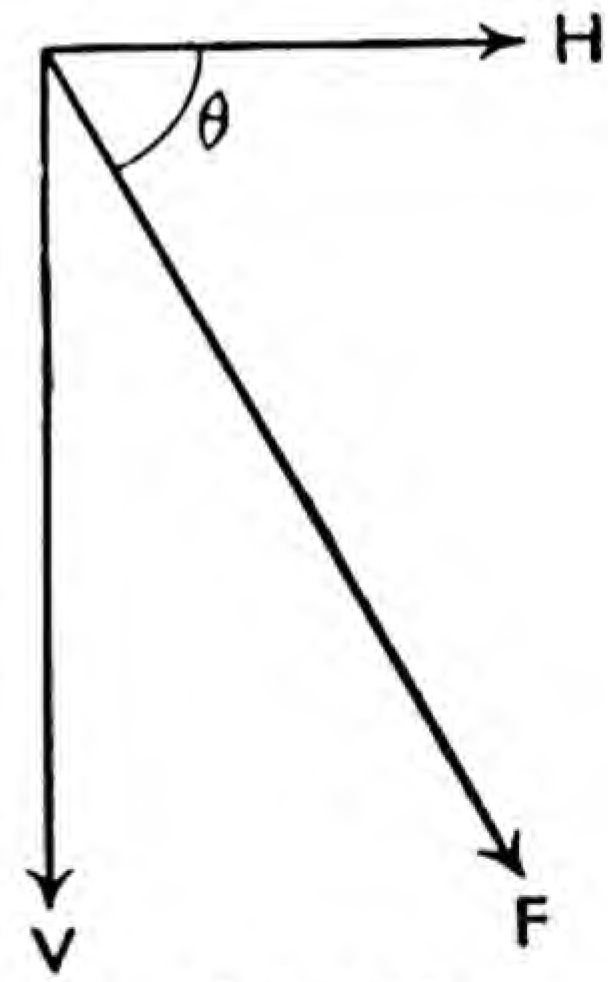


FIG. 105
Horizontal (H) and Vertical (V) Components of Earth's Magnetic Field (F) and Angle of Dip θ

The above-mentioned plane, in which a magnet completely free to swing sets itself, is called the *magnetic meridian.* It is normal to the Earth's surface and passes near the geographical poles. We shall consider the phenomena of terrestrial magnetism further in Chapter XV.

Unless special steps are taken to neutralize the effects of

the Earth's magnetization, all experiments and operations with magnets take place in a magnetic field, and this fact must be taken account of in all our observations. It is turned to good account in the mariner's compass and in other ways (see pp. 262–63).

Exploration of Magnetic Fields

A small magnet (usually called a *magnetic needle*) is often used to "explore" a magnetic field, for, if allowed to swing, it will lie along the direction of the line of force at the point where it is placed. If such a needle is placed in the Earth's field it will lie almost north and south, and if displaced from that position it will move back again, oscillating to and fro before coming to rest. If, now, a magnet is brought near, its field will modify the field in the neighbourhood of the needle, and the needle will take up a new position, which will be that of the resultant of the fields of the Earth and the magnet. This gives a simple means of measuring the pole-strength of a magnet (or the product of the pole strength and length * of the magnet—a more important quantity known as the *moment* of the magnet) if the strength of the Earth's field is known—or, if we do not know the strength of the Earth's field, of finding the ratio of the moments of two magnets. We have simply to place the magnets in turn in the same position with respect to the needle, and compare the deflections produced.

The deflections, of course, will vary with the position chosen for the magnets. We will consider only the two special positions which are of chief interest—namely, those in which the magnet, lying horizontally and perpendicular to the Earth's horizontal component, is placed (i) E. or W., and (ii) N. or S. of the needle. These positions are known as the *Tangent (A) and Tangent (B) Positions of Gauss.*

* By the "length" of a magnet is meant the distance between the poles. Since the poles are not quite at the ends of the bar, this distance is slightly less than the true length of the bar.

Tangent (A) and (B) Positions of Gauss

Let us take first the Tangent (A) position. This is illustrated in Fig. 106, where *ns* is the needle and NS the magnet.

FIG. 106
The Tangent (A) Position of Gauss

Consider the force on the pole n of the needle. If m is the pole-strength of the magnet, $2l$ its length, and r the distance from the nearer pole (shown N. in the figure) to any part of the very small needle, then a unit pole at n would be repelled by a force $\frac{m}{r^2}$, and if m' is the pole-strength of the needle, the repulsive force will therefore be $\frac{mm'}{r^2}$ in the direction of the axis of the magnet NS, since the whole of the needle may be regarded as lying on this axis. The pole S, distant $2l + r$ from n, will, however, attract n along the same line with the smaller force $\frac{mm'}{(2l + r)^2}$, and the resultant force on n due to the magnet will therefore be a repulsion P, given by

$$P = \frac{mm'}{r^2} - \frac{mm'}{(2l + r)^2} \quad . \quad (11{\cdot}1)$$

Clearly an equal force of attraction will act on the pole s, so that the needle will take up a position $n's'$ indicated in Fig. 107, in which the couple due to the field of the magnet will be balanced by that due to the Earth's horizontal field, H.

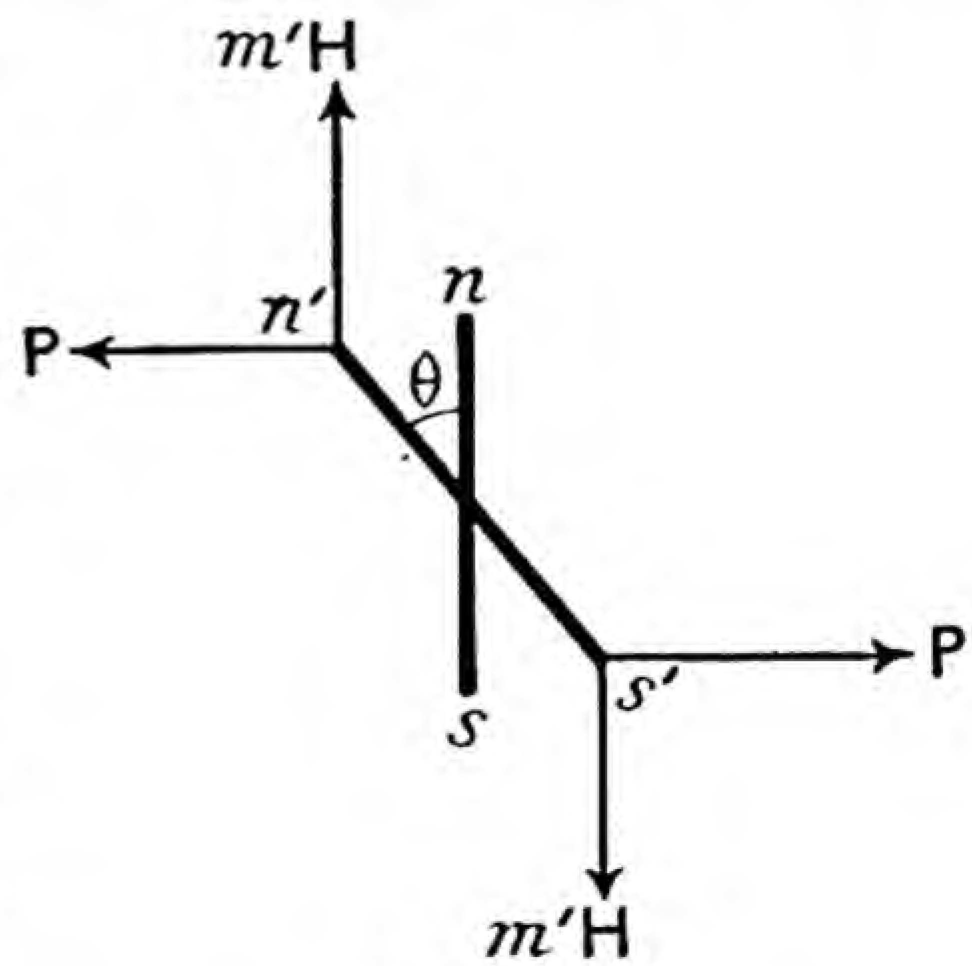


FIG. 107
Equilibrium Position $n's'$ of Needle in Tangent (A) Position of Gauss

If the angle of deflection is θ, the former couple is $Pd \cos \theta$, where d is the length of the needle ; and the latter couple is $m'Hd \sin \theta$, since $m'H$ is the force on each pole due to the Earth's field. We have therefore

$$mm'\left[\frac{1}{r^2} - \frac{1}{(2l+r)^2}\right] \cos \theta = m'H \sin \theta \quad . \quad (11\cdot2)$$

which reduces to $2lm \dfrac{2(l+r)}{r^2(2l+r)^2} = H \tan \theta \quad . \quad . \quad . \quad (11\cdot3)$

But $2lm$ is the moment of the magnet, M, say.

$$\text{Hence} \quad \frac{M}{H} = \frac{r^2(2l+r)^2}{2(l+r)} \tan \theta \quad . \quad . \quad . \quad (11\cdot4)$$

or, if the distance r is great compared with the length of the magnet, $2l$,

$$\frac{M}{H} = \frac{r^3}{2} \tan \theta \text{ approximately} \quad . \quad . \quad . \quad (11\cdot5)$$

Let us now take the Tangent (B) position, represented in Fig. 108. Here the force on s is an attraction towards N of $\dfrac{mm'}{\mathrm{N}s^2} = \dfrac{mm'}{l^2 + r^2}$ (r now being the distance of the needle from the *centre* of the magnet), combined with a repulsion from S of the same amount. The resultant P of these two forces is clearly parallel to the magnet, and if ϕ is the angle between the direction of either force and the original direction of the needle, the resultant is

$$P = 2\frac{mm'}{l^2 + r^2} \sin \phi \quad . \quad . \quad . \quad . \quad (11\cdot6)$$

There will be an equal force on n (the length of the needle again being negligible), and the needle will take up the position of equilibrium shown in Fig. 109, in which the couple

Fig. 108
The Tangent (B) Position of Gauss

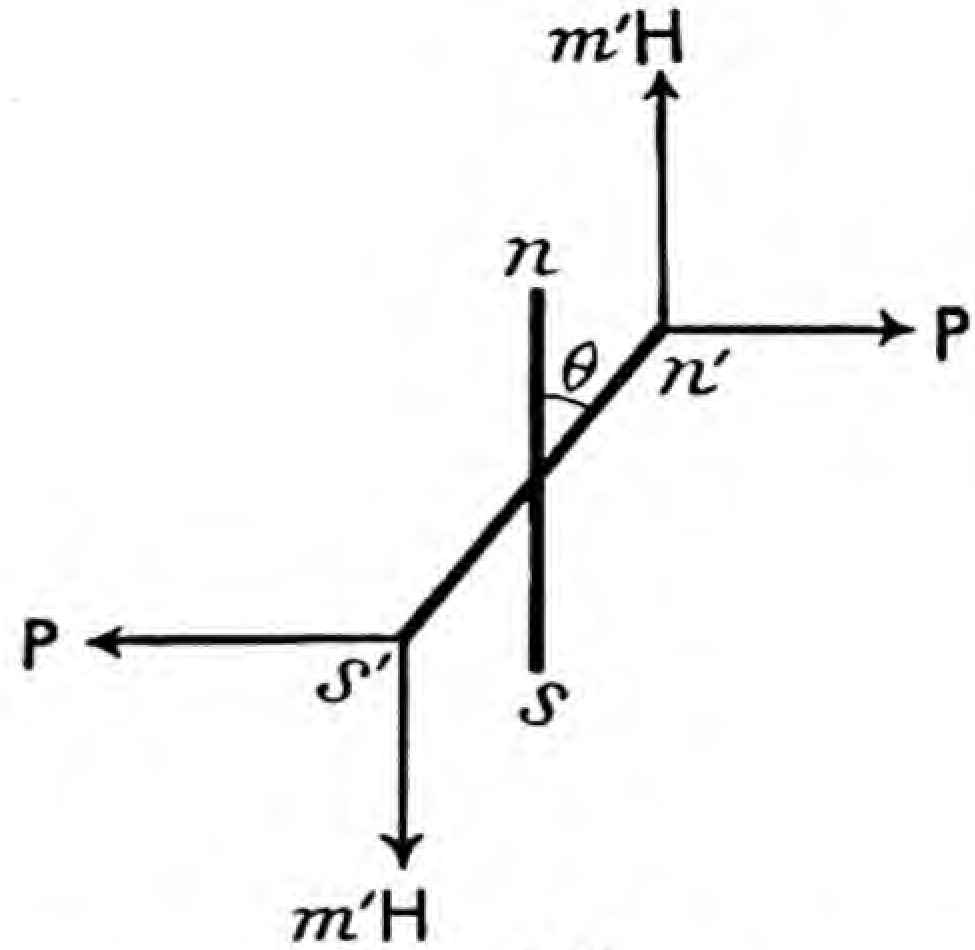


Fig. 109
Equilibrium position $n's'$ of Needle in Tangent (B) Position of Gauss

due to the deflecting forces will be equal to that due to the Earth's field. We shall have, then, as before,

$$Pd \cos \theta = m'Hd \sin \theta \quad . \quad . \quad . \quad (11{\cdot}7)$$

which now becomes

$$\frac{2\,mm' \sin \phi}{l^2 + r^2} d \cos \theta = m'Hd \sin \theta \quad . \quad . \quad (11{\cdot}8)$$

Now, from the figure,

$$\sin \phi = \frac{l}{\sqrt{l^2 + r^2}}$$

$$\text{so that} \quad \frac{2\,ml}{(l^2 + r^2)^{\frac{3}{2}}} = H \tan \theta \quad . \quad . \quad . \quad (11{\cdot}9)$$

$$\textit{i.e.} \quad \frac{M}{H} = (l^2 + r^2)^{\frac{3}{2}} \tan \theta \quad . \quad . \quad . \quad (11{\cdot}10)$$

If, as before, we may neglect l^2 in comparison with r^2, this becomes

$$\frac{M}{H} = r^3 \tan \theta \quad . \quad . \quad . \quad . \quad (11{\cdot}11)$$

The method of comparing the moments of two magnets is now obvious. Place them in turn in either the A or the B position, at the same distance r from the needle, and the moments of the magnets are then in the ratio of the tangents of the angles of deflection. If the size of either of the magnets is such that l cannot be safely neglected in comparison with r, this is no longer true, and we must use formula (11·4) or (11·10), in which l may have different values for the two magnets.

Oscillation Method of Comparing Magnetic Moments

There is another very convenient method of comparing the moments of two magnets, which requires no auxiliary needle. If a magnet is suspended at its centre so that it can move freely in a horizontal plane, it will, of course, set itself along the lines of horizontal force of the Earth's field. If now it is rotated horizontally through a small angle and then released, it will return towards its original position and oscillate to and fro for some time before coming to rest. It can be proved that the time t of each complete to and fro oscillation is independent of the amplitude so long as that is fairly small, and is given by

$$t = 2\pi\sqrt{\frac{I}{MH}} \quad . \quad . \quad . \quad . \quad (11{\cdot}12)$$

where I and M are the moment of inertia and magnetic moment of the magnet, respectively, and H is the horizontal component of the Earth's field. If we know both I and H we can therefore determine M by observing the time taken to make n oscillations (where, for accuracy, n is as large as possible) and dividing by n. This gives t, whereupon M can be calculated. If we do not know H we can still get the

ratio of the moments M_1 and M_2 of two magnets, for if we observe t for each we have clearly

$$\frac{t_1}{t_2} = \sqrt{\frac{I_1 M_2}{I_2 M_1}} \quad . \quad . \quad . \quad . \quad . \quad (11\cdot13)$$

$$\text{whence} \quad \frac{M_2}{M_1} = \frac{I_2}{I_1} \frac{t_1^{\,2}}{t_2^{\,2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11\cdot14)$$

Magnetizability of Materials

We have seen that when a piece of iron is placed in a magnetic field it sets itself in the direction of the lines of force, and, further, the lines of force crowd into the iron (see Fig. 103) so that more of them cross each sq. cm. of cross-section of the iron than crossed the same space before the iron was introduced. Now this means that the iron becomes equivalent to a magnet, for we know that the strength of the field at any place is proportional to the number of lines of force crossing unit area there (see p. 186), so that the force near the ends of the iron is increased exactly as though those ends were magnetic poles. The iron AB in Fig. 103 is therefore made into a magnet by being placed in the field of the magnet NS. If NS be removed, then AB may revert to its original non-magnetized state if the field is very weak, but if it is of considerable strength, AB retains some of its magnetism, though not all.

Investigation of Magnetic Properties: It is very important to examine the effect of a magnetic field on a piece of iron, steel, or other material, so that we may use it to produce the results most desired in given circumstances. To do this we place the material in a magnetic field, the strength of which we can vary at will, and observe how the magnetization changes as the field is increased or decreased. The most convenient way of producing the field is to make use of the fact that an electric current in a closed circuit behaves as a magnet, *i.e.* produces a magnetic field. We therefore

prepare a circuit of the form shown in Fig. 110, in which a portion AB has the form of a number of windings with a space inside in which a cylindrical rod can be inserted. Such a portion, in which the windings are very close together and may be carried forward and back a number of times so as to form several layers, is known as a *solenoid.** It acts as a magnet, and if, on looking in the direction AB, the direction of the current is clockwise, then B will act as a N. and A as a S. pole. If the direction of the current is reversed, then the polarity is reversed. The strength of the field inside the solenoid is proportional to the number of turns of wire per unit length and to the current; if these are respectively n and C, then the strength of the field in the centre of the solenoid (the regions near the ends being excluded) is uniform and equal to $4\pi n \frac{C}{10}$ when C is measured in amperes (see p. 215). We can thus, by varying C, produce whatever field we wish in the solenoid, and we do this by including a variable resistance (known as a *rheostat*) in the circuit. There is, of course, an upper limit to C, fixed by the resistance of the remainder of the circuit and the E.M.F. available, but the range below this can be made great enough for our purpose.

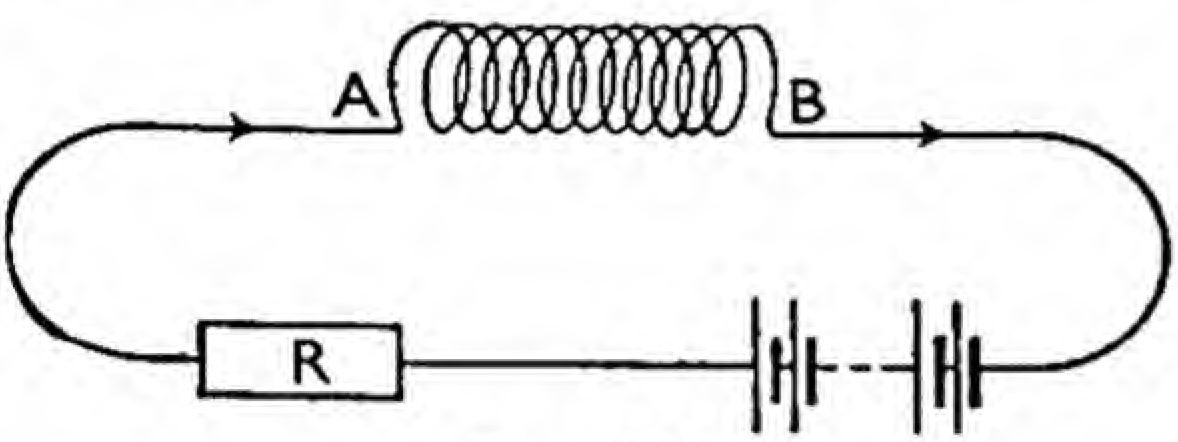


FIG. 110
Circuit containing Solenoid AB

Suppose, now, a current flows in the circuit. Then the solenoid acts as a magnet and the lines of magnetic force take the form shown in Fig. 111. If a bar of iron (or other magnetizable material; we will speak of iron for simplicity)

* Whenever different portions of a circuit are placed in contact, or very close together, as in such windings, they must be insulated from one another, or a " short circuit " will be formed. This is usually done by covering the wires by some non-conducting material such as cotton or silk.

is placed along the axis of the solenoid, the lines of force crowd into it and the solenoid and iron form a much stronger

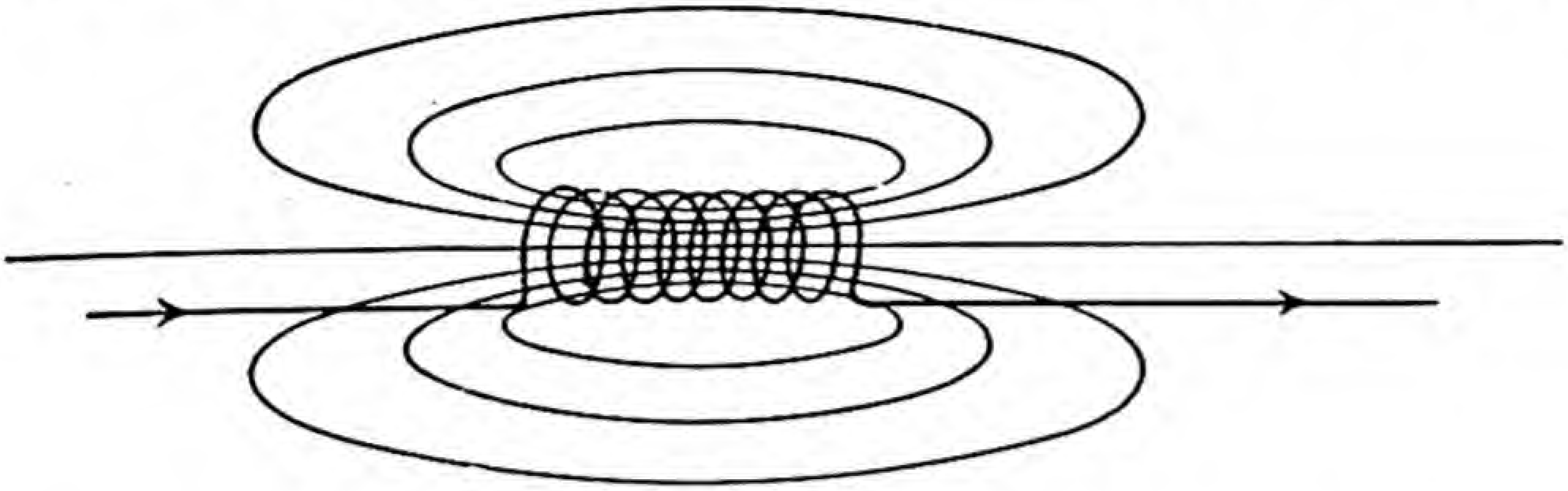


FIG. 111
Lines of Magnetic Force in Field of Solenoid carrying Current

magnet, with lines of force more numerous but otherwise arranged as before. The increase in strength of the field may be attributed to the magnetization of the iron. Such a magnet is known as an *electromagnet*, and is always used when strong magnetic fields are required. The iron is often bent round so that the poles face one another (see Fig. 112); the magnet is then known as a *horse-shoe* electromagnet. The field between the poles of such a magnet can be made very strong.

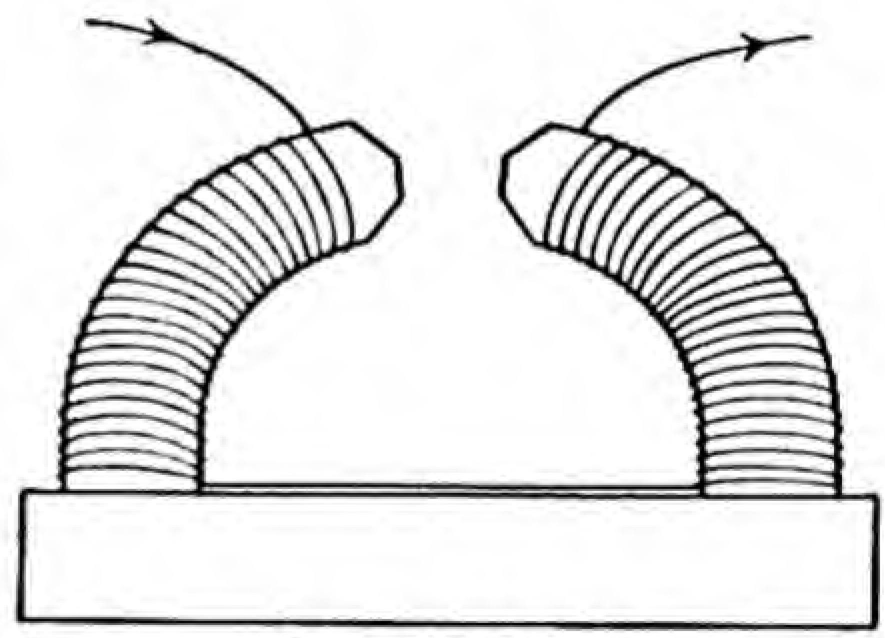


FIG. 112
Horseshoe Electromagnet

Characterization of Magnetizability: We must now explain some terms that are used in describing this process of magnetization. The field in the solenoid before the iron is introduced is called the *inducing field*; its strength F, as we have said, is equal to $\frac{2}{5}\pi nC$. In this field, F lines of force cross each unit of area normal to the axis of the solenoid. When the iron is inserted, a greater number of lines of force cross each unit of area in the iron. This number, B, is called the *induction*. The introduction of the iron has thus caused an increase of $B - F$ in the number of lines per sq. cm.,

and this increase we attribute to the magnetization of the iron.

Now suppose the cross-sectional area of the iron is a. Then the number of lines of force issuing from each pole on account of the magnetization of the iron is $a(B - F)$. But if the pole-strength of the iron is m, this number of lines of force (see p. 186) is $4\pi m$.

$$\text{Hence} \quad 4\pi m = a(B - F) \quad . \quad . \quad . \quad . \quad (11\cdot15)$$

$$\text{so that} \quad \frac{m}{a} = \frac{1}{4\pi}(B - F) \quad . \quad . \quad . \quad (11\cdot16)$$

The quantity $\frac{m}{a}$, which is the pole-strength per unit area, is called the *intensity of magnetization* of the iron, and is usually denoted by I. We have therefore

$$B = F + 4\pi I \quad . \quad . \quad . \quad . \quad (11\cdot17)$$

This equation expresses the fact that the field B in the iron is equal to the original inducing field F plus the field $4\pi I$ arising from the magnetization of the iron. Although, in order to form a clear picture of what is happening, we have supposed the field F to be produced by an electric current, that is not necessary; the equation is perfectly general. Thus, if F is the Earth's field H, and a piece of iron is placed along the lines of force of this field, then the intensity of magnetization produced will be related to the crowding of lines of force into the iron by the same equation.

The ratio $\frac{B}{F}$ is called the *permeability* (μ) of the iron, and the ratio $\frac{I}{F}$ is called the *susceptibility* (κ). Equation (11·17) thus gives

$$\mu = 1 + 4\pi\kappa \quad . \quad . \quad . \quad . \quad . \quad (11\cdot18)$$

Clearly μ and κ increase or decrease together; they are really different ways of expressing the same general property of magnetizability of the material.

Measurement of Permeability

We may now see how the permeability of a substance can be measured. Fig. 113 shows two exactly similar solenoids, AB and CD, placed in a straight line in the same circuit, and wound in such a way that when a current is flowing, B and C become similar poles. Their effects therefore neutralize one another at a point midway between B and C, and a magnetic needle placed there will therefore lie in the direction of the Earth's field, just as though there

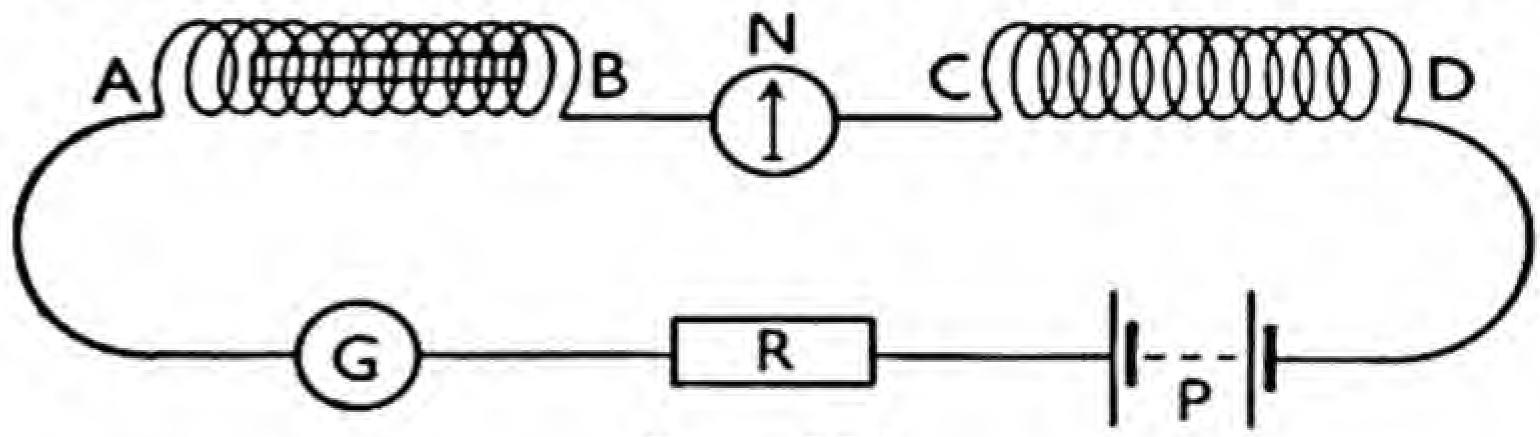


FIG. 113

Apparatus for measuring Permeability

AB, CD Similar Solenoids — G Galvanometer
N Compass Needle — R Resistance
P Battery

were no current in the solenoids. Let the common axis of the solenoids lie " magnetic east and west " (*i.e.* perpendicular to the magnetic meridian) ; the needle will then be at right angles to the axis. (To correct for unavoidable dissimilarities in the solenoids, the needle is placed where it actually lies in this direction, whether or not it is then exactly midway between B and C.) Now place a long thin rod of magnetizable material inside and along the axis of AB, the rod being not long enough to reach to either end. When a current flows it becomes magnetized, and the needle is deflected through an angle θ, say, solely on account of the magnetization of the rod, since the effect of the inducing field F is neutralized by the similar field arising from CD. Now this is the Tangent (A) position of Gauss. Hence, if M is the moment of the magnetized rod,

$$\frac{M}{H} = \frac{r^2(2l + r)^2}{2(l + r)} \tan \theta \qquad . \quad . \quad . \quad (11{\cdot}4)$$

where $2l$ is the length of the rod and r the distance of its nearer end from the needle. But $M = 2lm = 2laI$.

$$\text{Hence } I = \frac{Hr^2(2l + r)^2}{4la(l + r)} \tan \theta. \quad . \quad . \quad (11{\cdot}19)$$

The inducing field F is $\frac{2}{5} \pi nC$ (p. 196). Hence the induction B is

$$B = \frac{2}{5} \pi nC + \frac{\pi Hr^2(2l + r)^2}{la(l + r)} \tan \theta \ . \quad . \quad . \quad (11{\cdot}20)$$

from (11·17), and the permeability μ is given by

$$\mu = 1 + \frac{5Hr^2(2l + r)^2}{2nCla(l + r)} \tan \theta. \quad . \quad . \quad (11{\cdot}21)$$

The Hysteresis Curve

Let us now see how the magnetization of the rod varies with the inducing field. Let us start with no current in the solenoids, and an unmagnetized rod. The needle will then be undeflected. Now let a small current C_1 pass through the solenoids. We shall then have $F = \frac{2}{5} \pi nC_1$, while B has the value obtained by putting $C = C_1$ in (11·20), and taking the observed value of θ. Gradually increase the current, at each step noting the values of C and θ and calculating the corresponding values of F and B; then plot B against F, as in Fig. 114. The curve will be found to take the form Oa, the rate of increase of B at first showing an acceleration but afterwards slowing down until an almost constant value of B is reached, when further increase of the current (and therefore of F) makes scarcely any difference to B.

At this point let us gradually reduce the current, observing the deflection and calculating F and B at each step. It will be found that the curve does not retrace its path, but that for each value of F the value of B is higher on the return

than on the initial journey, so that when the current is reduced to zero again the needle still remains deflected, corresponding to a value Ob of B. This means that the iron retains some of its magnetization after the magnetizing field has been completely removed ; it could be taken out and used as a permanent magnet. If, however, we do not remove it, but slowly increase the current *in the opposite direction*, thus creating a field opposed to the original one, we find that the curve

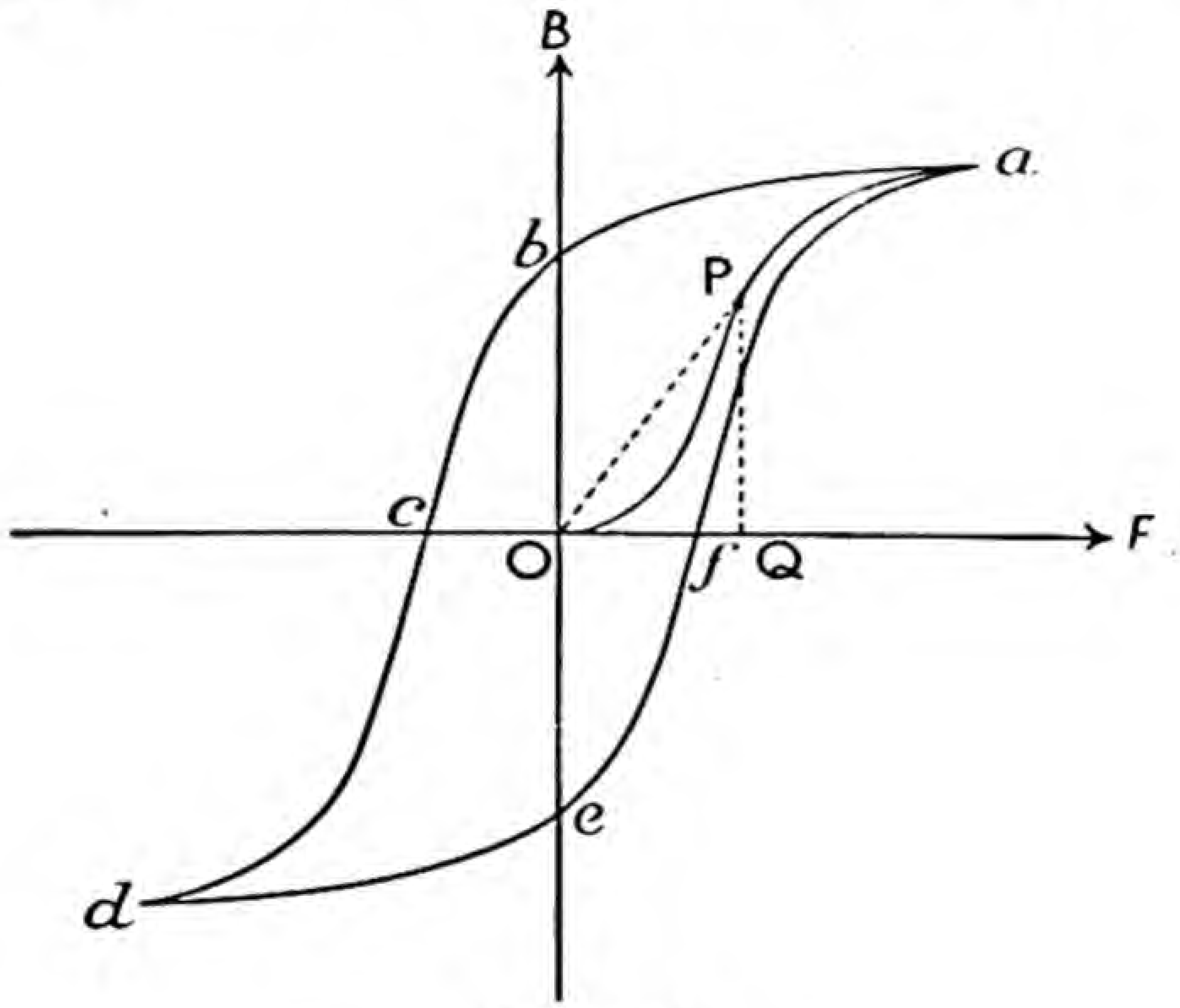


FIG. 114
Hysteresis Curve

continues along bc to d, where the induction and the intensity of magnetization again reach a maximum value, but of opposite polarity to the first. If at this point we reduce the current again back to zero and then increase it in the original direction up to its first maximum, we find that the curve follows the course *defa*.

A curve of this general type is shown by all magnetizable materials, but the details of its form vary widely from one specimen to another. For some the figure is long and thin ; for others it is short and stout. It is called a *hysteresis* curve,

and is of great importance because it shows what magnetic behaviour we may expect from the material which it characterizes. The distance Ob (*i.e.* the magnetization remaining when the inducing field is removed) is called the ***retentivity*** of the material ; it measures the power of retaining magnetization. Oc is called the *coercivity* ; it shows the force necessary to demagnetize the specimen. The permeability $\left(\frac{B}{F}\right)$ is clearly variable, for at any point such as P it is equal to $\frac{PQ}{OQ}$, *i.e.* $\tan POQ$, and this ratio varies with the position of P on the curve. It increases at first, reaches a maximum value, and then decreases again.

A great deal of work has been done on the magnetic properties of alloys, and it is possible now to go a long way towards making an alloy to suit any particular specification —*e.g.* high retentivity, small coercivity, etc. Alloys of much higher permeability than iron can be made, particularly certain cobalt steels.

The effect of temperature on magnetic characteristics is important. It appears that as the temperature of a specimen is raised, its permeability for very weak inducing fields is increased, while that for larger fields is decreased. There is a critical temperature (785° C. for iron) at which the substance ceases to be magnetic altogether.

Explanation of the Hysteresis Curve : If we remember that each atom or molecule of a magnetizable substance is a magnet, we can get a simple explanation of the hysteresis curve. In an unmagnetized specimen the atoms have no particular orientation, and their fields on the whole neutralize one another. When an inducing field is applied, however, they tend to set themselves along its lines of force. With weak fields the alignment is only partial, so that the specimen shows a weak magnetization, but when the inducing field reaches a certain value the atomic magnets swing over more or less

uniformly into its direction. We then have a rapid increase of B, after which little more can be done, for when the atomic magnets are perfectly aligned along the field the intensity of magnetization is as great as is possible for the specimen. We thus reach the point a, beyond which there is scarcely any increase of B. When the inducing field is removed, the atomic magnets to a large extent keep their alignment, since only the atomic or molecular motions which constitute their heat tend to remove them from it. Thus the magnetization is largely retained. At high temperatures the heat motions may be such as to prevent any permanent alignment altogether and thus prevent magnetization.

Diamagnetism

Only a few substances show strong magnetizability, but almost every substance is magnetizable to some extent. The intense magnetization shown by iron, cobalt, nickel, and some alloys is known as *ferromagnetism*. The same type of magnetization shown weakly by other substances is called *paramagnetism*, and a magnetization of the reverse kind (*i.e.* one in which the substance is repelled from the poles of a ferromagnet, sets itself perpendicular to a ferromagnetic field, and moves from stronger to weaker parts of such a field) is called *diamagnetism*. There is reason to believe that all substances are diamagnetic to a very slight extent, though in only a few (bismuth is the best) is the property easily made manifest. In para- and ferromagnets the paramagnetism is superposed on and overpowers the diamagnetism. Paramagnetism and diamagnetism are thus not alternatives, like positive and negative electrification, between which a body, so to speak, must choose. They are believed to originate independently in the atomic and molecular structure of matter, though ultimately they are both associated with the electrical character of the elementary particles.

EXERCISES

1. Draw the lines of force in the space surrounding two magnets of equal dimensions and pole-strength, placed a slight distance apart in the same straight line, with (*a*) the N. pole of each on the left, and (*b*) the N. pole of one and the S. pole of the other on the left.

2. You are given two rods of steel, exactly similar in appearance, and are told that one is magnetized and the other unmagnetized. How could you determine, without any auxiliary apparatus whatever, which rod is magnetized ?

3. Describe and explain a method of shielding a space from the action of external magnetic fields.

4. What reasons are there for believing that the Earth is a magnet ? Describe the general character of terrestrial magnetism.

5. A bar magnet is placed in the A Position of Gauss with its centre 15 cm. from a compass needle, and the deflection produced is reduced to zero by a second bar magnet placed in the B Position of Gauss with its centre 20 cm. from the needle. If the lengths of the magnets are respectively 10 and 12 cm., find the ratio of their magnetic moments.

6. Compare the magnetic fields at two places where a suspended bar magnet makes 20 and 30 oscillations a minute respectively.

7. Explain the terms *inducing field*, *induction*, *intensity of magnetization*, *intensity of magnetic field*, *permeability*, and *susceptibility*. What relations exist between the numerical measures of the properties denoted by them ?

8. Describe a method of obtaining the hysteresis curve for a specimen of iron. What information can be obtained from such a curve ?

CHAPTER XII

ELECTROMAGNETISM

It has already been explained that the magnetic effects considered in the last chapter, though having no visible connection with electricity, are believed to arise as a consequence of the electrical nature of the ultimate particles of matter. The electrons moving in atoms constitute small electric currents, and a magnetic field is believed to be inseparably associated with an electric current. It has been proved experimentally, for example, that not only electric currents from batteries, but also ordinary "static" electric charges when set in rapid motion are accompanied by magnetic fields; the moving charges, in fact, behave in all respects like an electric current, so that we are quite justified in attributing magnetic properties to atoms containing revolving electrons. In this chapter we shall see how to determine the character and strength of the magnetic field arising from a given current of electricity.

Magnetic Field due to Current

Let us begin with the simplest case—a long straight wire carrying a uniform current. By experiments with small magnetic needles or iron filings it can be shown that the space surrounding such a wire is a magnetic field of which the lines of force are circles whose centres lie on the wire. Thus, if, as in Fig. 115, the wire be passed perpendicularly through a card on which iron filings are distributed, the filings set themselves along circles as shown. To find the direction of the field we must use a needle. We then find that if we look along the wire in the direction of the current, the N. pole

tends to move in a clockwise direction, so that the field is clockwise.

The strength of the field may be found from what is known as *Laplace's Law*, which is as follows. Consider a short length dl of the wire. The intensity, dF, of the magnetic field due

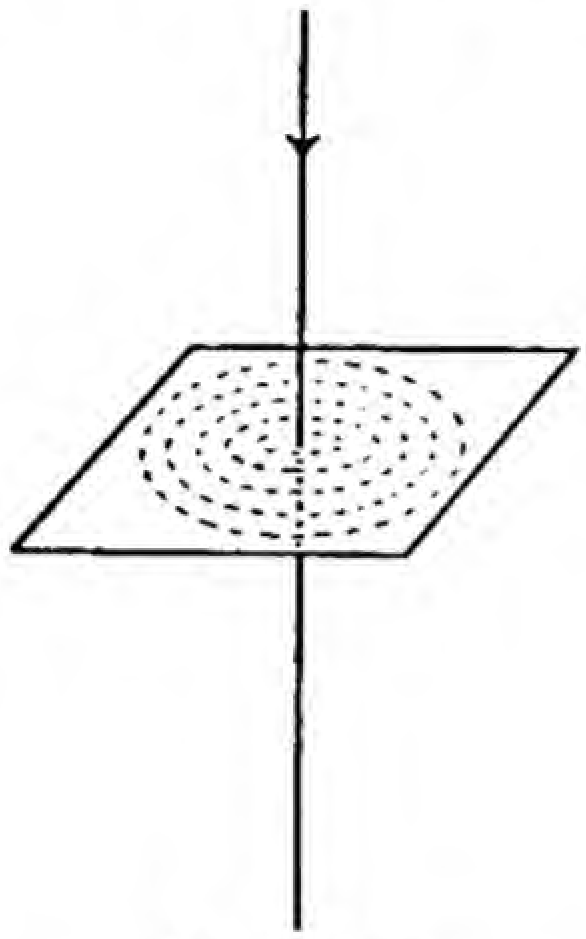


FIG. 115
Lines of Force in Field surrounding Linear Current

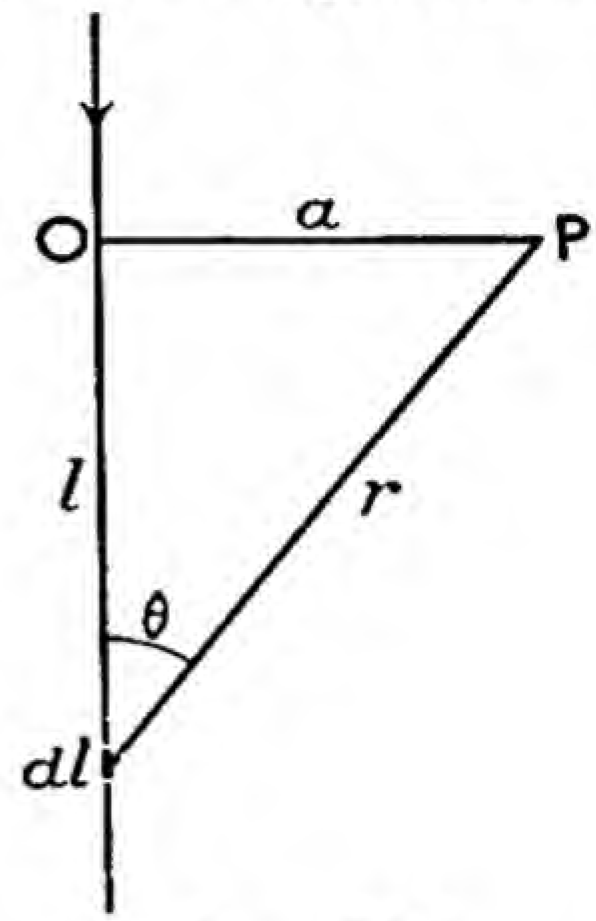


FIG. 116
Diagram illustrating calculation of Field at P due to Element dl of Linear Current

to a current C in this length, at a point P distant r from dl, is directly proportional to C, to dl, and to $\sin\theta$ (where θ is the angle between dl and the direction of P), and inversely proportional to r^2 (see Fig. 116). We may therefore write:

$$dF = k\,\frac{Cdl\sin\theta}{r^2} \quad . \quad . \quad . \quad . \quad (12{\cdot}1)$$

where k is a constant. The total field F at P is obtained by integrating this expression along the whole length of the wire, thus:

$$F = kC\int\frac{dl\sin\theta}{r^2} \quad . \quad . \quad . \quad . \quad (12{\cdot}2)$$

Now if PO ($=a$) is the perpendicular from P to the wire, and the distance from O to dl is l. then $r = a\operatorname{cosec}\theta$ and

$l = a \cot \theta$, so that $dl = - a \operatorname{cosec}^2 \theta d\theta$ by simple differentiation. Further, if the wire is long compared with a, we may take the limits of θ as 0 and π. We have then

$$F = - kC\int_{\pi}^{0} \frac{\sin \theta}{a}\, d\theta = \frac{2kC}{a} \quad . \quad . \quad . \quad (12{\cdot}3)$$

The resultant field is therefore directly proportional to the current and inversely proportional to the distance of P from the wire.

Circular Current : This result is useful when we are considering the immediate neighbourhood of the wire, for

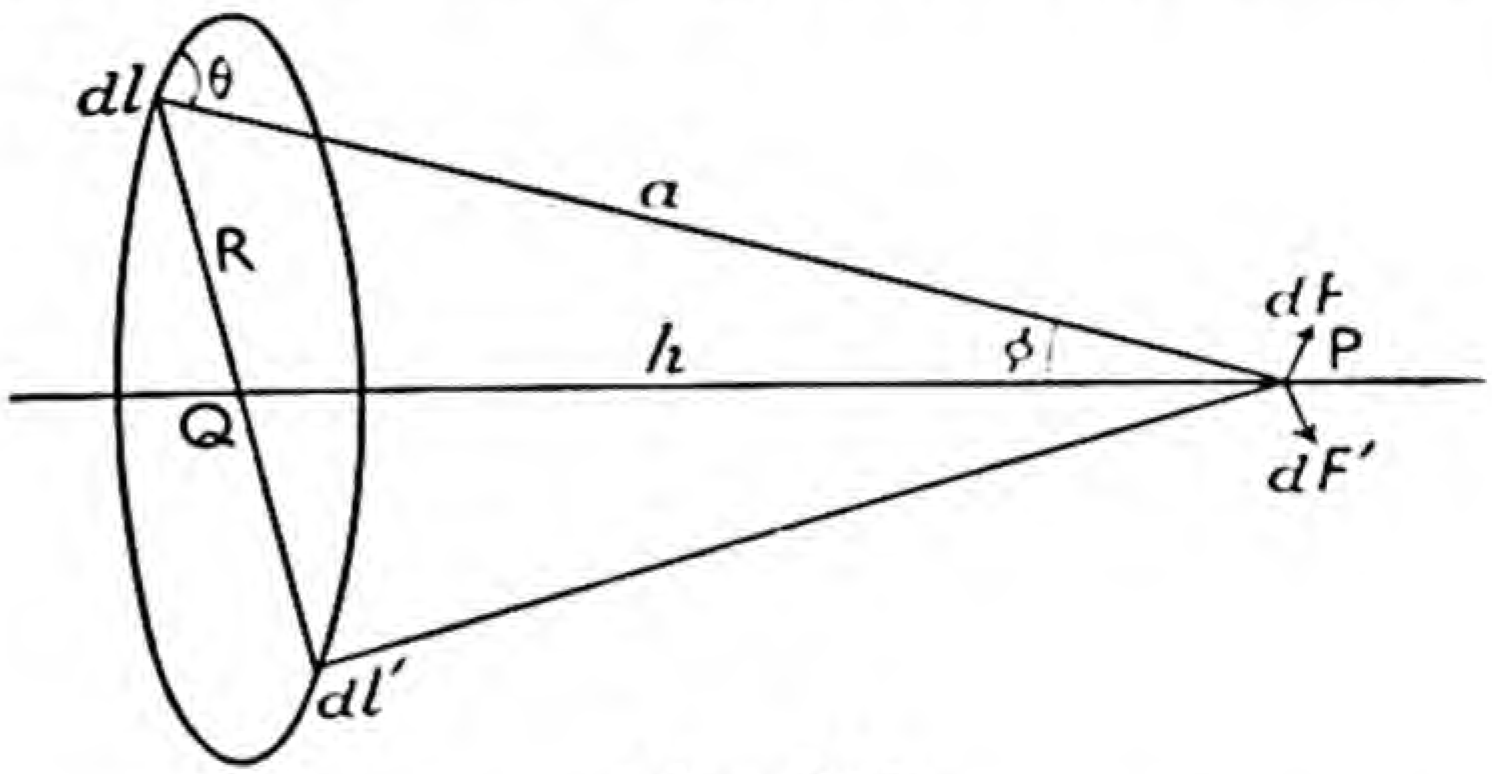


FIG. 117

Diagram illustrating calculation of Field on Axis of Circular Current

clearly it is only when a is small that we can regard θ as extending from 0 to π, and ignore, as we have done, the effect of the rest of the circuit containing the wire. In the more general case we must take the whole circuit into account, and we shall now consider a circular current and calculate the magnetic field at a point on its axis. In this case (see Fig. 117) θ is clearly always $\frac{\pi}{2}$ for all points of the circuit, and r and a become identical. We have therefore

$$dF = k\,\frac{Cdl}{a^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12{\cdot}4)$$

from (12·1). Now the contributions to the field by the various

elements dl of the circuit will be in different directions, for they will be perpendicular to the lines joining dl to P. It is easily seen, however, that if we resolve the force dF into two components, $dF \sin \phi$ along QP and $dF \cos \phi$ perpendicular to QP, the latter component will, when the whole circuit is taken into account, amount to nothing, since, for every element dl there will be an opposite element dl' for which the component perpendicular to QP will neutralize that due to dl. We need, therefore, consider only the component $dF \sin \phi$ in the integration, and the resultant direction of the field will be along the axis QP, either from Q towards P or P towards Q, according to the direction of the current. We thus obtain

$$F = \frac{kC}{a^2}\int \sin \phi \, dl \quad . \quad . \quad . \quad . \quad . \quad (12\cdot5)$$

Now $\sin \phi$ is constant all round the wire, and equal to $\frac{R}{a}$, where R is the radius of the circuit; and also, since we integrate over the whole circuit, the limits of l are 0 and $2\pi R$. Hence

$$F = \frac{kCR}{a^3}\int_0^{2\pi R} dl = k\frac{2\pi R^2 C}{a^3} = k\frac{2\pi R^2 C}{(h^2 + R^2)^{\frac{3}{2}}} \quad . \quad (12\cdot6)$$

where $h = $ QP.

If the circuit has n turns of wire instead of one, the total length is multiplied by n; the limits of integration are then 0 and $2\pi Rn$ and the field strength is

$$F = k\frac{2\pi n R^2 C}{(h^2+R^2)^{\frac{3}{2}}} \quad . \quad . \quad . \quad . \quad . \quad (12\cdot7)$$

It is therefore proportional to the current and to the number of turns of wire, and it is easy to see that it decreases as h or R increases. For a given coil, therefore, it is greatest at the centre Q, where its value, since $h = 0$, is

$$F = k\frac{2\pi n C}{R} \quad . \quad . \quad . \quad . \quad . \quad (12\cdot8)$$

The Electromagnetic Unit of Current

We may use this result to determine the constant k. For, if we set the coil in the magnetic meridian and place a magnetic needle at its centre, the needle will be deflected until it is in equilibrium between the couples due to the Earth's field and to the field produced by the coil. Equating the moments of these couples, we can determine F. If then we count n, measure R, and know the strength of the current C, we can solve (12·8) for k. It is found that when C is measured in electrostatic units of electric charge per second, k is equal to about $\frac{1}{3\times 10^{10}}$. It is inconvenient to have this very small number burdening our equations, and so, since our interest in electric currents is mostly concerned with their magnetic effects, we choose a new unit, the *electromagnetic* unit, which is 3×10^{10} times as large as the electrostatic unit. We can easily obtain a formal definition of this unit from (12·8). For, if we take $R = 1$ cm., $n = 1$, and $C = 1$ electromagnetic unit, we have, for the field at the centre of the coil,

$$F = \frac{1}{3\times 10^{10}} \frac{2\pi\times 3\times 10^{10}}{1} = 2\pi \text{ dynes} \quad . \quad (12{\cdot}9)$$

Hence *the electromagnetic unit of current is that current which, when flowing in a circular coil of 1 cm. radius and one turn of wire, produces a magnetic field of 2π dynes per unit pole at the centre.* It is easily seen that if we measure C in electromagnetic (e.m.) units, k is equal to unity and may be omitted from the equations.

The Electromagnetic System of Units

The e.m. unit of current is but one of a whole system of units—the *electromagnetic system* (also sometimes called the C.G.S. system)—which is substituted for the electrostatic (e.s.) system when we are dealing with currents of electricity instead of static electric charges. If we regard the e.m. unit of current as the passage of 1 e.m. unit of electricity per second, then it obviously follows that 1 e.m. unit of charge, or quantity

of electricity, is 3×10^{10} times 1 e.s. unit. In most problems there is little danger of confusion between the units because the e.m. system (or a modification of it called the *practical* system, to be described presently) is almost invariably used in problems concerning currents of electricity, and the e.s. system in problems concerning static charges. We have to be careful, however, when we are considering atomic problems, for here we regard moving electrons as currents, and if we measure the charge of an electron in e.s. units, the current consisting of a number of electrons passing a given point per second will come out in e.s. units. The number 3×10^{10}—the ratio of the units—is generally denoted by the letter c, and is equal to the velocity of light in C.G.S. units. This is an important factor in the *electromagnetic theory of light*, according to which light consists of electromagnetic waves; this theory, however, is beyond our scope.

There are e.m. units of P.D. and resistance also. Two points have a P.D. of one e.m. unit if one erg of work is done when 1 e.m. unit of electricity moves from the point of higher to the point of lower potential. Since the e.s. unit of P.D. is defined similarly, but with 1 e.s. instead of 1 e.m. unit of quantity, it follows that the e.m. unit of P.D. is $\frac{1}{c}$ times the e.s. unit, since, for the same amount of work, the charge must be inversely proportional to the P.D.

The unit of resistance is derived from Ohm's law. If we have a circuit carrying a current, the potential of course decreases in the direction of the current. If the current is 1 e.m. unit, and we choose two points on the circuit whose P.D. is 1 e.m. unit, the resistance between those points will be 1 e.m. unit.

The Practical System of Units

It may be thought unnecessary to have two independent systems to deal with, but the duplication is justified because e.s. forces and e.m. forces are of such greatly different orders

of magnitude that a common system would entail the introduction of inconveniently large or small numbers into many of our calculations. Even two systems are insufficient to overcome the difficulty altogether, because if we adopt the e.m. system to obtain a convenient measure of current strength, we find we have an inconveniently small measure of P.D., so a third system of units—the *practical* system—is introduced in order to combine the best aspects of the other two. In this system the unit of electric current is the *ampere*, which is simply $\frac{1}{10}$ of the e.m. unit, and the corresponding unit of quantity is the *coulomb* ; an ampere is, therefore, a current of one coulomb per second. We have already mentioned the coulomb (p. 166) in connection with electrolysis. The practical unit of P.D., or E.M.F., is the *volt* (10^8 e.m. units), and the practical unit of resistance is the *ohm*, which, to satisfy Ohm's law, must be $\frac{10^8}{10^{-1}} = 10^9$ e.m. units. It is found to be equal to the resistance of a column of mercury 106·3 cm. long and 1 sq. mm. in cross-section, at 0° C.

The following table gives a comparison of the values of the chief units in the three systems, in terms of the e.m. units :

	e.m.	e.s.	Practical
Quantity	1	$\frac{1}{c}$	$\frac{1}{10}$ = 1 coulomb
Current	1	$\frac{1}{c}$	$\frac{1}{10}$ = 1 ampere
P.D. or E.M.F. . .	1	c	10^8 = 1 volt
Resistance	1	c^2	10^9 = 1 ohm

There are, of course, corresponding relations for all other electrical and magnetic measurements—*e.g.* capacity, the practical unit of which is called the *farad*—which may be

obtained from this table and the definitions of the quantities. One important magnitude is the amount of work done when a unit charge falls through unit P.D. In both the e.m. and the e.s. systems this is 1 erg since the unit of P.D. is derived by making it so. In the practical system, however, we see that 1 coulomb $\times$ 1 volt $= 10^7$ e.m. units, *i.e.* 10^7 ergs. This is called the *joule*. Hence in measuring the heating effect of a current, for example (see p. 178), if the current C and P.D. E are measured in amperes and volts, respectively, the energy generated in time t seconds is ECt joules.*

Magnetic Moment of Circular Current

We may now return to the magnetic effects of a current. A circular coil such as that considered on p. 207 is clearly equivalent to a very short magnet, for the lines of force are perpendicular to the plane of the coil, and the two faces of the coil may be regarded as the N. and S. poles (Fig. 118). Such a coil has a magnetic moment, but we cannot calculate its value from the ordinary definition (pole-strength $\times$ distance between poles), because the distance between the "poles" is indefinite. We can, however, calculate it as follows:

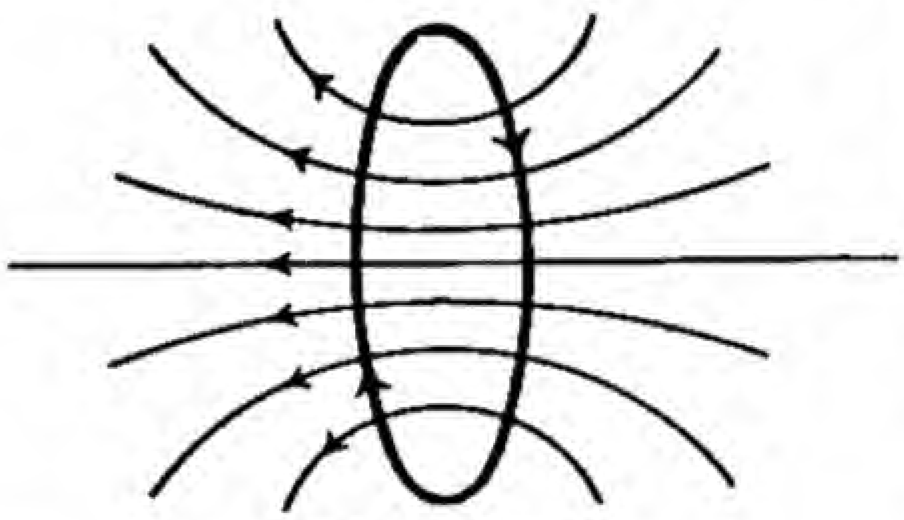


FIG. 118
Lines of Force due to Circular Current, showing its Resemblance to a Magnet

On p. 191 we found that the force on a unit N-pole (*i.e.* the intensity of the field) at a point on the axis of a magnet was

$$P = m\left[\frac{1}{r^2} - \frac{1}{(2l+r)^2}\right] = M\frac{2(l+r)}{r^2(2l+r)^2} \quad . \quad (12\cdot10)$$

* The whole question of electrical units is now under revision. A new system (the "M.K.S." system, in which the metre, kilogram and second are the units of length, mass and time, respectively, and the electrical units are chosen with a view of making the equations of electromagnetism as simple as possible) is now being widely used. In the present unsettled state, however, it has been thought best here to keep to the traditional systems of units.

(obtained by putting m′ = 1 in (11·1)), and if we choose a point so far from the magnet that $2l$ is negligible compared with r, this becomes

$$P = \frac{2M}{r^3} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12\cdot11)$$

Now we have found that the intensity of the field on the axis of a circular coil of one turn, carrying a current of C e.m. units, is

$$F = \frac{2\pi R^2 C}{h^3} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12\cdot12)$$

(obtained by putting $k = 1$, $n = 1$, and ignoring R^2 in comparison with h^2 in (12·7)), if again we choose a point so far from the coil that R is negligible compared with h. Hence the magnet and coil produce the same field at distant axial points if $F = P$; *i.e.* since r is equivalent to h, if

$$M = \pi R^2 C \quad . \quad . \quad . \quad . \quad . \quad . \quad (12\cdot13)$$

We may, therefore, *so far as the effect at such points is concerned*, treat the coil as a magnet of moment equal to $\pi R^2 C$, and define the magnetic moment of a circular current as *the product of the current strength* (*in e.m. units*) *and the area of the circuit.* If the coil has n turns, the magnetic moment is, of course, n times as great.

Magnetic Field on Axis of Solenoid

We may now calculate the magnetic field on the axis of a solenoid. Suppose that, including all the layers there might be, there are n turns of wire in each unit of length of the solenoid. Then, in a small length, dh, there will be ndh turns, and if dh is small enough, we may regard the centres of these turns as a single point, Q, on the axis (Fig. 119), distant h from P, the point at which the field is to be found. We may then apply (12·7), putting $k = 1$ (*i.e.* measuring C in e.m. units) and putting ndh for n, when we obtain

$$dF = \frac{2\pi n dh R^2 C}{(h^2 + R^2)^{\frac{3}{2}}} \quad . \quad . \quad . \quad . \quad . \quad (12\cdot14)$$

giving the contribution of the length dh to the field at P. The field due to the whole solenoid is therefore obtained by

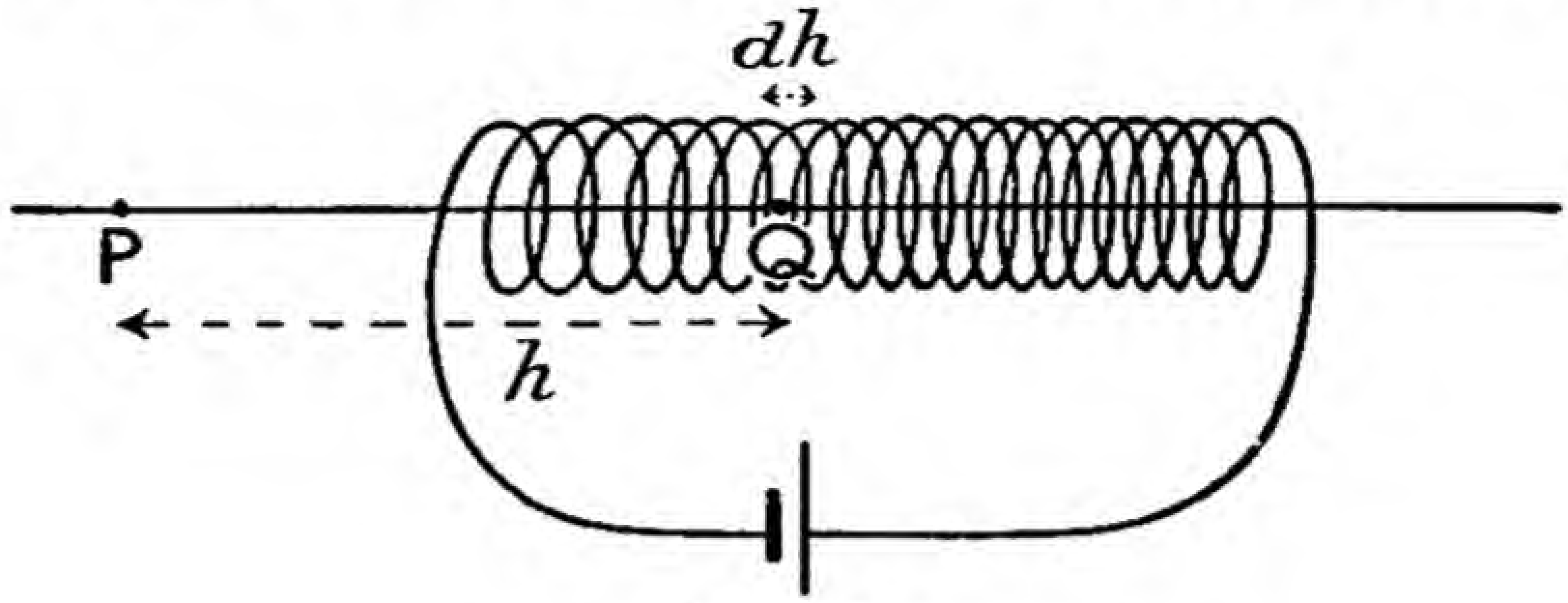


FIG. 119
Solenoid carrying Current

integrating (12·14) along the whole solenoid. If h_1 and h_2 are the distances of P from the two ends, we have

$$F = 2\pi nR^2C\int_{h_1}^{h_2} \frac{dh}{(h^2+R^2)^{\frac{3}{2}}}$$

$$= 2\pi nC\left[\frac{h_2}{\sqrt{h_2^2+R^2}} - \frac{h_1}{\sqrt{h_1^2+R^2}}\right] \quad . \quad (12·15)$$

In practice, solenoids are usually long and thin, so that R is small compared with their length. Two cases are then of special interest: (i) when P is at or near one end; and (ii) when P is inside the solenoid at a considerable distance from either end (*e.g.* at Q in Fig. 119).

In the first case, h_1 say, is zero, while h_2 is h, the length of the solenoid. Neglecting R in comparison with h we then see that

$$F = 2\pi nC \quad . \quad . \quad . \quad . \quad . \quad (12·16)$$

In the second case, similarly neglecting R, we have apparently $F = 2\pi nC(1 - 1) = 0$. But now the two ends are on opposite sides of P, so that h_1 and h_2 must be given opposite signs. The field is therefore

$$F = 2\pi nC(1 + 1) = 4\pi nC \quad . \quad . \quad . \quad (12·17)$$

At any axial point within the solenoid, then, whose distance from either end is large compared with the radius, the field is approximately uniform and equal to $4\pi nC$. If the solenoid contains a core of permeability μ, the induction is then of course

$$B = 4\pi nC\mu \quad . \quad . \quad . \quad . \quad . \quad (12\cdot18)$$

If the current C is measured in amperes, we must divide the measure by 10 to get its value in e.m. units. In that case $\mathrm{F} = \frac{2}{5}\pi nC$. This is the formula we used on p. 196.

Electromagnetic Induction

Since an electric current creates a magnetic field in the surrounding space, it is natural to inquire whether, if a closed conducting circuit be placed in a magnetic field, a current

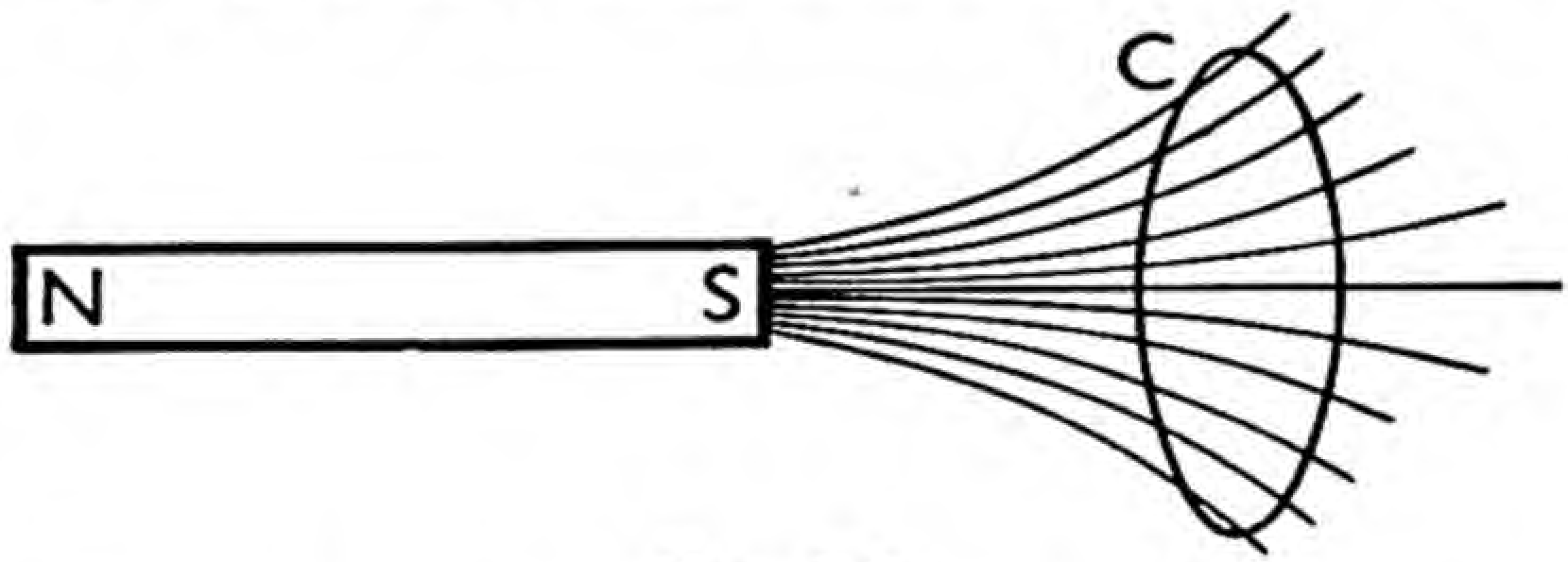


FIG. 120
Magnetic Flux through Coil C due to Magnet NS

will be created in it. The answer is that it will if the circuit is moving or the field is varying in such a way that the number of lines of magnetic force passing through the circuit is changing, but not otherwise. For example, in Fig. 120, so long as the magnet NS and the coil C do not move with respect to one another, no current will flow through the coil, but if they approach or recede from one another, a current will start and will exist as long as the number of magnetic lines of force through the coil is increasing or diminishing. Again, in Fig. 121, if the magnetic field is in

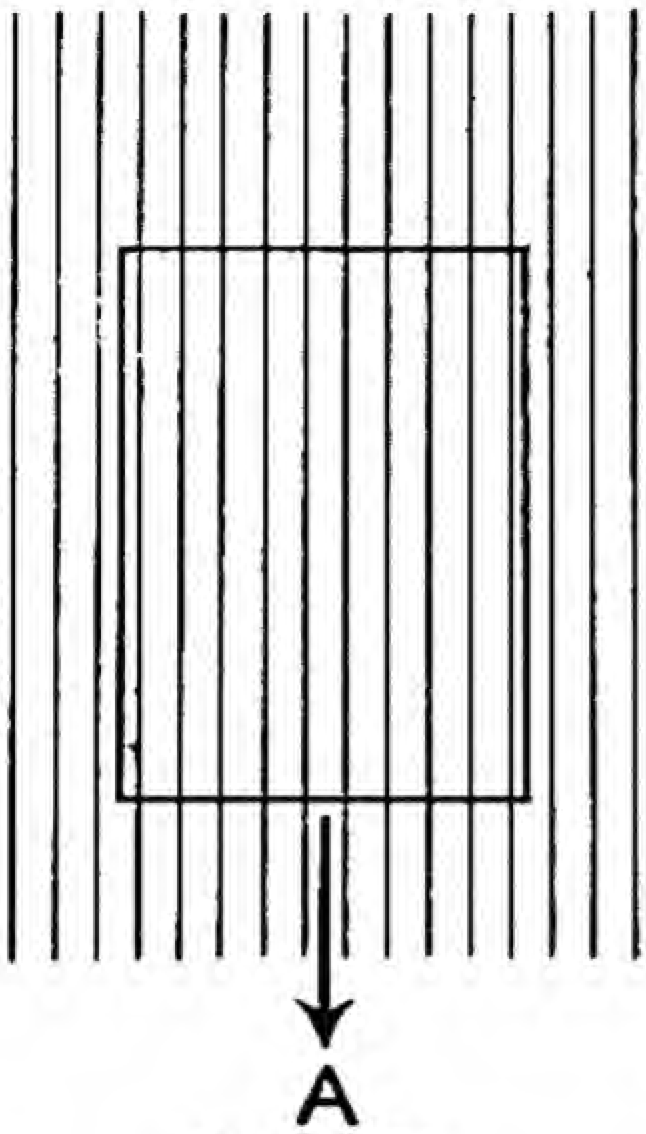


FIG. 121
Rectangular Coil moving parallel to Lines of Force of Uniform Magnetic Field

the direction A, then a movement of the rectangular circuit in that direction will not give rise to a current, but if the circuit is rotated about a line in the plane of the paper parallel or perpendicular to A, the number of lines of force passing through it will grow, and a current will flow.

Furthermore, not only will a current be created in a circuit so moving in a magnetic field, but also a circuit carrying a current, when placed in a magnetic field, will in general tend to move. The rule is that the movement will be such as to enable the circuit to encompass as many lines of magnetic force as possible. Thus, in Fig. 122, if the field is in the direction A, the circuit will set itself perpendicular to A, for in that position the maximum number of lines of force will pass through it.

The Right- and Left-hand Rules

These simple facts (known as *electromagnetic induction*) form the basis of every electric motor and dynamo on which we now depend for many of the amenities of civilized life. Consider a closed circuit in a magnetic field. If the circuit moves as described above, a current is created in it, and we have, in principle, a dynamo. If, on the other hand, a current is started in it, then it tends to move, and we have a motor. Useful rules have been given by Fleming for determining the directions of current and motion respectively in the two cases. Let the thumb, first

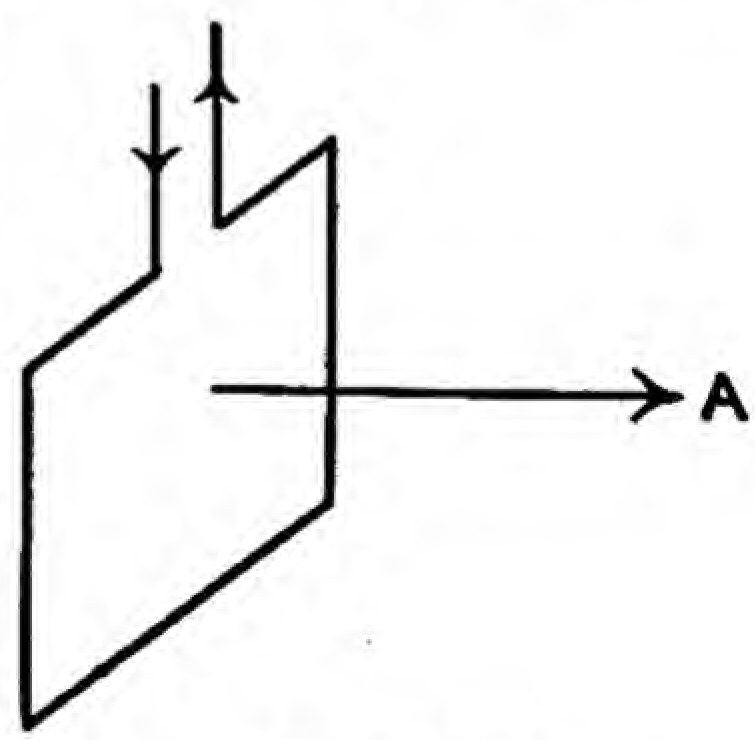


FIG. 122
Orientation of Coil carrying Current in Magnetic Field A

finger and second finger of the hand be placed so that they are mutually perpendicular, and let the thumb represent

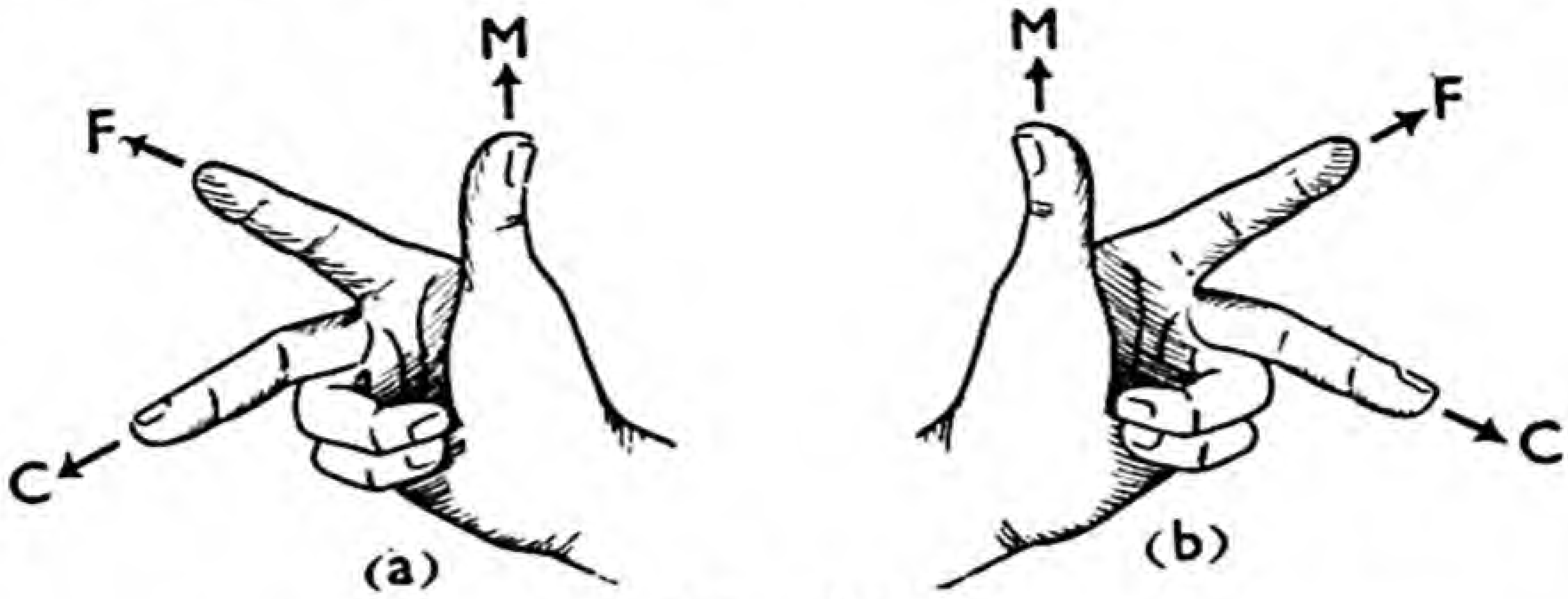


FIG. 123

Directions of (*a*) Induced Current (right hand) and (*b*) Induced Motion (left hand) in Magnetic Field

F Direction of Field C Direction of Current M Direction of Motion

the direction of motion, the first finger the direction of the magnetic field, and the second finger the direction of the current. Then the direction in which a current flows when the circuit is moved (the case of the dynamo) is given by the direction of the second finger of the *right* hand; while the direction of motion if a current is established in the circuit (the case of the motor) is given by the direction of the thumb of the *left* hand (Fig. 123).

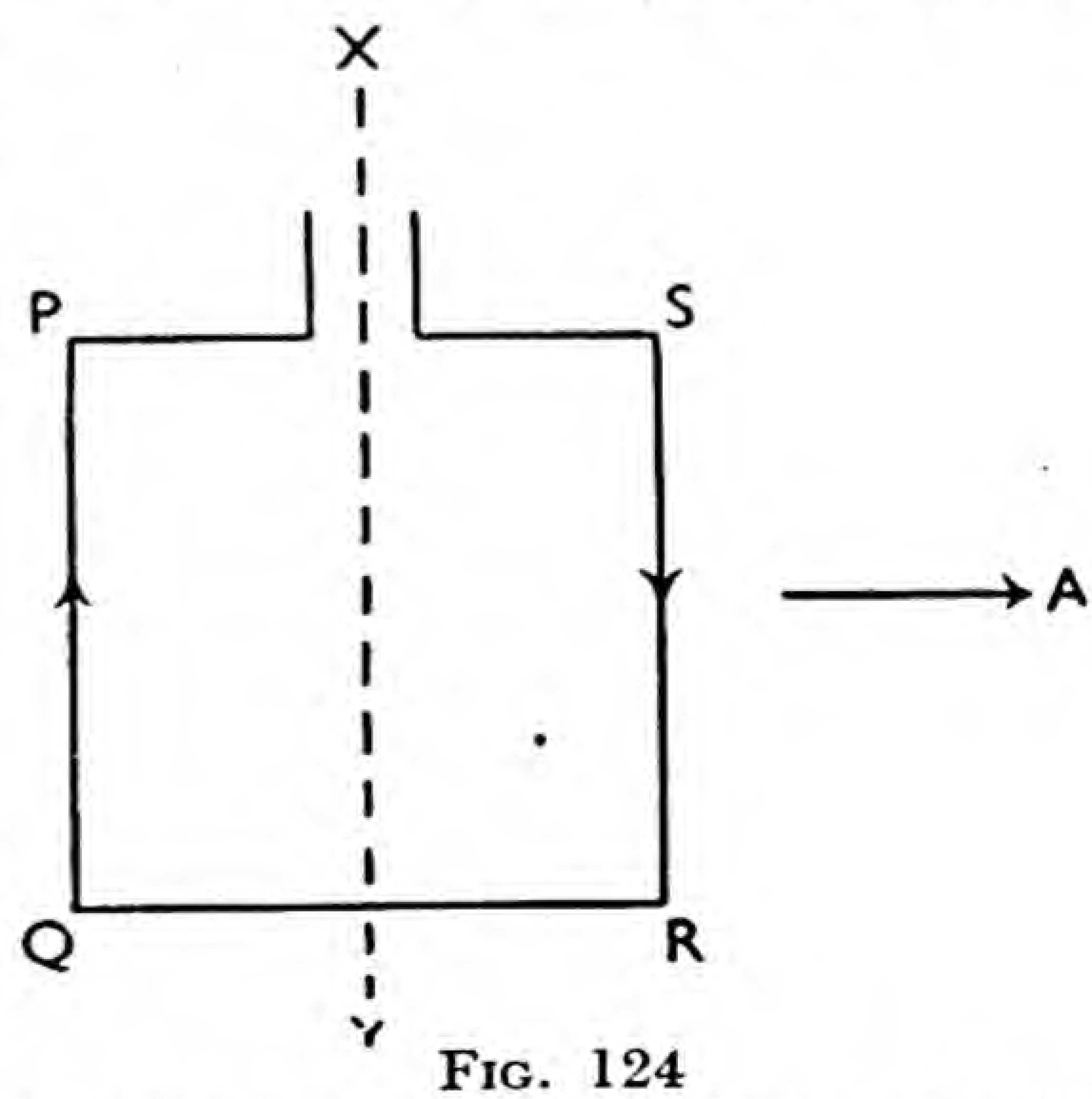


FIG. 124

Rectangular Current in Plane Parallel to Magnetic Field A

As an example, suppose we have a rectangular circuit placed so that its plane lies in the direction of a magnetic field A (Fig. 124), and let a current flow through it in the direction shown by the arrows. Applying the

left-hand rule, we see that the side SR will move upwards from the paper, and the side PQ downwards ; *i.e.* the coil will revolve about the axis XY. The revolution will continue until the coil is perpendicular to the plane of the paper, and there (possibly after a few to-and-fro oscillations) it will come to rest, for the tendency is always to bring SR as far as possible above, and PQ as far as possible below, the plane of the paper. This is clearly the position in which the coil embraces the maximum number of lines of force, so that by Fleming's rule we have arrived back at the condition mentioned earlier.

Eddy Currents

Another interesting and important example is the production of currents in a sheet of conducting material which moves relatively to a magnetic field. Suppose a magnet is suspended horizontally over the centre of a horizontal copper disc, and the disc is rotated about a vertical axis through its centre (Fig. 125). It is found that the magnet, if free to do so, rotates also in the same direction, though less rapidly. The number of lines of magnetic force through a small element of area A of the disc varies as the rotation brings the element to different parts of the field, and a current is therefore created round that element. This in its turn creates a magnetic field which reacts on the magnet, causing it to rotate in a direction which, by Lenz's law (see p. 220) tends to oppose the creation of the induced current. This is clearly the same direction as that in which the disc rotates, for if both rotated

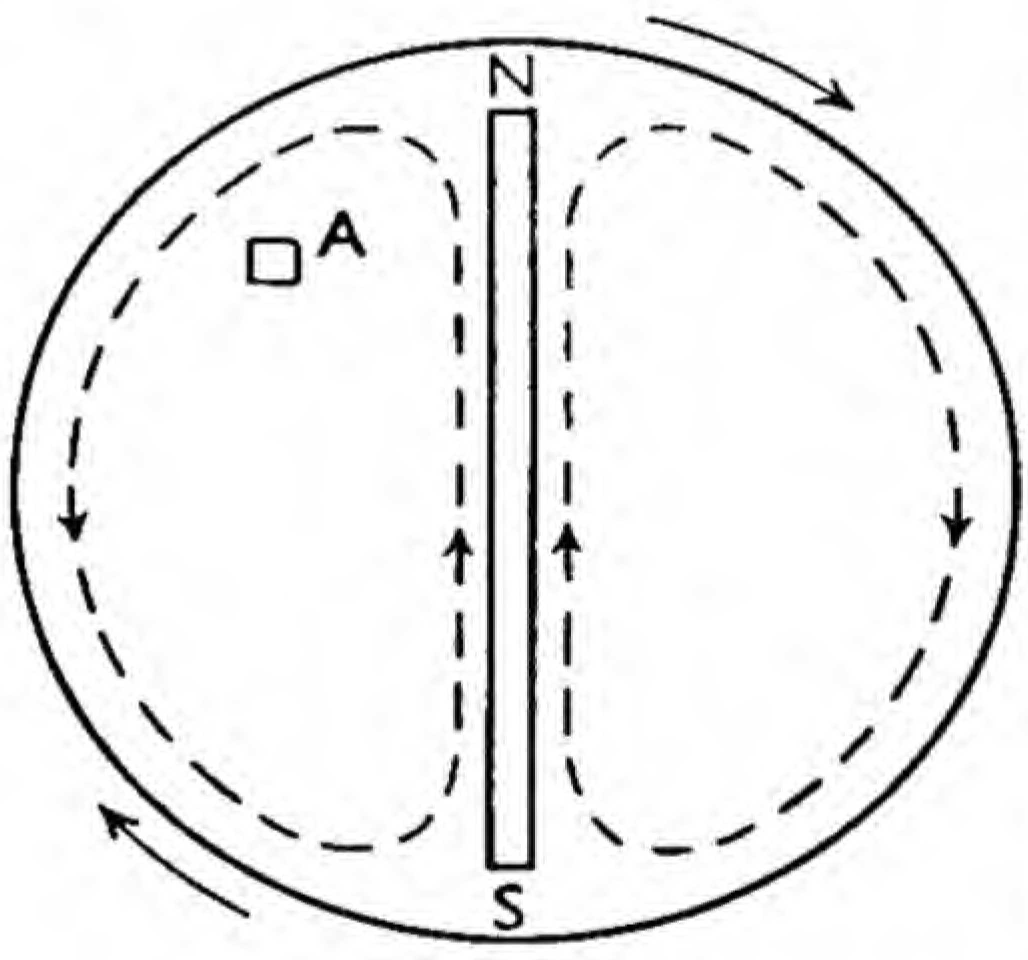


Fig. 125
Eddy Currents in Conducting Sheet rotating in Magnetic Field

in the same direction at the same rate there would be no relative motion between them, and so no induced current.

Every element of area of the disc behaves in the same way, and it can be shown that the resultant current forms two closed curves, following the courses shown by the dotted lines in Fig. 125. These currents are called *eddy* or *Foucault* currents, and their effects are often very striking, and can sometimes be turned to good account. They can be demonstrated very effectively by passing a thin copper sheet edgewise between the poles of a powerful horseshoe electromagnet. A strong resistance is experienced, and the feeling is as though one were cutting cheese rather than air. The eddy currents set up in the sheet tend, in the magnetic field, to move it backwards, to oppose the motion which gives rise to them. The energy expended in overcoming this resistance appears as heat in the sheet.

Mutual Induction

A very important consequence of the facts we are considering is the phenomenon known as *mutual induction.* Suppose we have two circuits, A and B (Fig. 126), and suppose a steady current flows in A. A magnetic field then surrounds A, and its lines of force pass through B, but so long as the current in A is constant, no current will flow in B, for there is no variation in the number of lines of force passing through that circuit. But suppose the current through A changes. Then a current will immediately start in B and continue as long as the change continues. In particular, a current will always flow in B when the current starts in A and when it is stopped. For, in the first case, the current

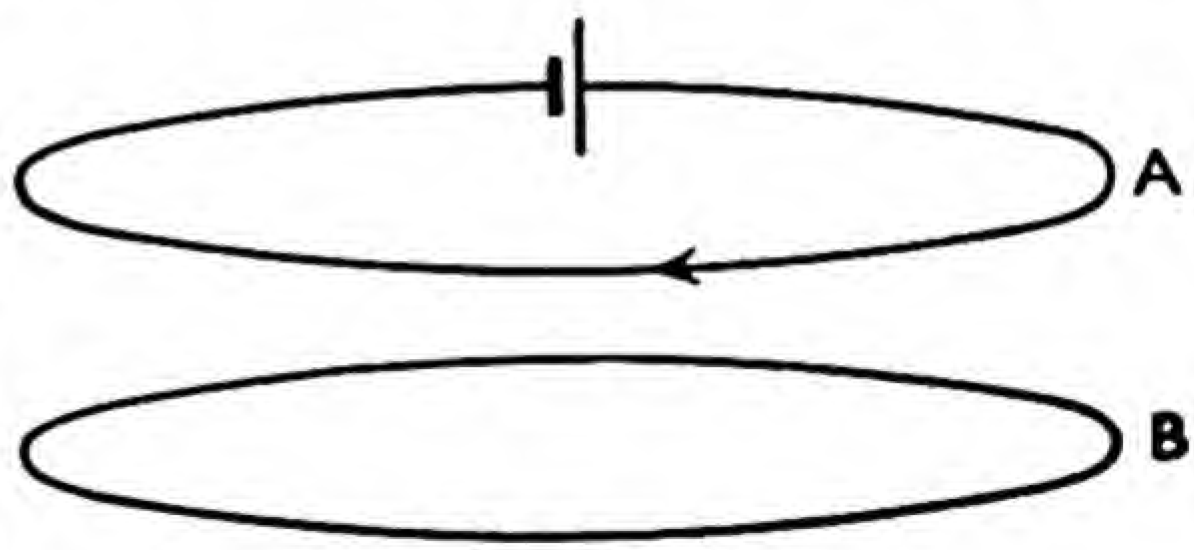


FIG. 126
Coils arranged to show Mutual Induction

must rise from zero to its steady value, and during its growth the field through B will grow ; and, in the second case, the current must fall from its steady value to zero, again causing a change in the field through B.

Direction of Current : The direction of the current in B in all cases is given by a law known as *Lenz's Law*, which states that the direction will be such as to oppose the change which causes it. Thus, when the current is growing in A, the magnetic field through B is increasing. The current in B will then be in such a direction as to create an opposite field, so as to retard this growth. With the configuration shown in Fig. 126, we see that the current in B will then be in the opposite direction to that in A. On the other hand, when the current in A is decreasing, a current in the *same* direction will exist in B, so as to tend to maintain the former value of the field.

Lenz's law applies to all cases of electromagnetic induction, whether two circuits are involved or not. In every case the current or movement which is induced is such as to oppose the cause of the induction.

Magnitude of Current : The magnitude of the current will, of course, depend on the resistance of the circuit, and it is better, therefore, to think of the magnitude of the E.M.F. set up, which is independent of the resistance. It can be shown that this, in e.m. units, is numerically equal to the rate at which the number of lines of magnetic force passing through the circuit is changing. It will be remembered that if the strength of the magnetic field at the position of the coil is N, we say that N lines of force pass through a unit of area normal to their direction. Hence we may say that the E.M.F. set up in the circuit is equal to $-\frac{dN}{dt}$ (the minus sign indicating, in accordance with Lenz's law, that the E.M.F. opposes the change in N), and if this is one line of force per second, then unit E.M.F. is set up.

This is sometimes chosen as the means of *defining* the e.m. unit of E.M.F. It can be proved that the unit so defined is equal to the P.D. between two points, such that one erg of work is done when one e.m. unit of electricity passes from one to the other. The definition is therefore consistent with that of the unit of P.D. given on p. 210.

Self Induction

When a current starts to flow in A (Fig. 126) a magnetic field begins to grow, and the number of lines of force passing through A, as well as the number through B, increases. An induced E.M.F. is therefore set up in A also, and by Lenz's law it is in the opposite direction to the E.M.F. which creates the original current. The effect is to delay the attainment of the final steady value of the current in A. If, when this value is attained, the current is stopped, then similarly an induced E.M.F. is generated which tends to make it continue, so that it does not stop abruptly but dies away more or less gradually. (If the circuit is suddenly broken the process is very quick, but not absolutely instantaneous.) This phenomenon is known as *self induction.*

Measurement of Self and Mutual Induction

Self and mutual induction are measured by two quantities known as *coefficients of self and mutual inductance* respectively. They are defined as follows. The coefficient of self inductance of a circuit is *the number of lines of magnetic force which are made to pass through the circuit by the steady flow of* 1 *e.m. unit of current through the circuit.* Similarly, the coefficient of mutual inductance of two circuits is *the number of lines of magnetic force which are made to pass through one circuit by the steady flow of* 1 *e.m. unit of current through the other.* It does not matter in which of the two circuits the current flows, provided that, if a circuit consists of several turns of wire, the total number of lines of force passing through it is reckoned as the sum of the numbers passing through all the turns ; thus, if m lines actually

pass through a coil containing n turns, the value of $\mathcal{N}$ adopted must be mn.

The coefficients of self and mutual inductance are generally represented by L and M respectively. Hence, if $\mathcal{N}_1$ and $\mathcal{N}_2$ represent the total magnetic *flux*, as it is called, through A and B respectively, we have

$$\mathcal{N}_1 = LC \quad . \quad . \quad . \quad . \quad . \quad (12{\cdot}19)$$

for a single circuit, and

$$\mathcal{N}_2 = MC \quad . \quad . \quad . \quad . \quad . \quad (12{\cdot}20)$$

for a pair of circuits. The corresponding E.M.F.'s set up by a variation of current, the circuits being otherwise unaltered, are accordingly

$$E_1 = -\frac{d\mathcal{N}_1}{dt} = -L\frac{dC}{dt} \quad . \quad . \quad . \quad (12{\cdot}21)$$

and $$E_2 = -\frac{d\mathcal{N}_2}{dt} = -M\frac{dC}{dt} \quad . \quad . \quad . \quad (12{\cdot}22)$$

It should be noticed that L and M depend not only on the circuits concerned, but on other circumstances as well. For instance, if a piece of iron be placed within A, $\mathcal{N}_1$, and possibly $\mathcal{N}_2$, is increased, and therefore L and M also.

Clearly the unit of self inductance is that of a circuit in which the passage of 1 e.m. unit of current creates a flux of one line of magnetic force. This is an inconvenient unit, however, and the practical unit is that of a circuit in which the passage of 1 ampere creates a flux of 10^9 lines of force. This is called the *henry*. It is used as a unit of mutual inductance also, the current then being in one circuit and the flux through the other. From equations (12·21) and (12·22) and the definitions already given, the student may easily verify that the self inductance (or mutual inductance, with obvious modifications) of a circuit is one henry when a variation of current of 1 ampere per second creates an E.M.F. of 1 volt.

Transformers

Equation (12·22) shows how we may obtain a high E.M.F. from a low one, or vice versa. Suppose we have two solenoids, one surrounding the other, as illustrated diagrammatically in Fig. 127, of which the inner has n_1 and the outer n_2 turns per unit length. If the current in the first (called the *primary*) is rapidly and repeatedly started and stopped, the magnetic flux through the other (the *secondary*) undergoes continuous oscillation, and an alternating E.M.F. is therefore set up. The number of lines of force due to a current C in the primary is (see p. 214)

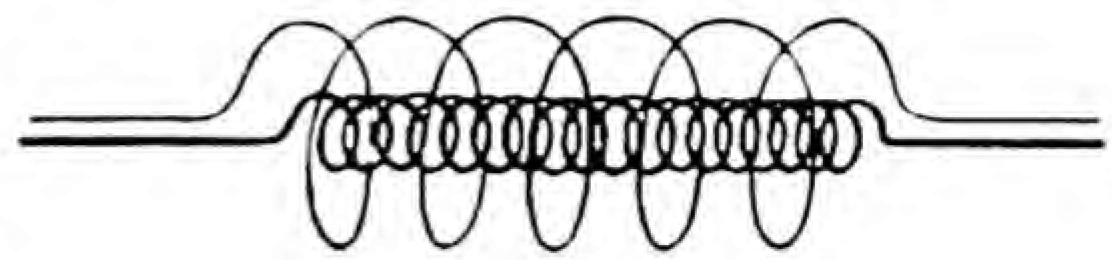


FIG. 127
Solenoids arranged to illustrate principle of Transformer

$$\mathcal{N}_1 = 4\pi n_1 CA \quad . \quad . \quad . \quad . \quad (12{\cdot}23)$$

where C is in e.m. units and A is the area of cross-section of the primary ; and since this number passes through each turn of the secondary, the flux $\mathcal{N}_2$ through the secondary (of length l, say) is $\mathcal{N}_1 n_2 l$, *i.e.*

$$\mathcal{N}_2 = 4\pi n_1 CA n_2 l \quad . \quad . \quad . \quad . \quad (12{\cdot}24)$$

Hence the E.M.F. set up in the secondary is

$$E = -\ 4\pi n_1 n_2 l A \frac{dC}{dt} \quad . \quad . \quad . \quad . \quad (12{\cdot}25)$$

The quantities n_1, n_2, lA, and $\frac{dC}{dt}$ in this expression are at our disposal, and we may choose them to obtain the result we desire. To get a large E.M.F., for instance, we may make rapid variations of current (choosing, let us say, n_1 small, so that the resistance of the primary may be small and its current large) and make n_2 large. A further increase is obtainable by placing iron of permeability μ inside the solenoids. In that case the E.M.F. given above is multiplied by μ.

This is the principle of the *transformer*, which is an instrument for obtaining a large from a small E.M.F. (a *step-up* transformer) or a small from a large E.M.F. (a *step-down* transformer). The solenoids just described constitute a transformer, and although in practice it may be necessary to modify their form somewhat, they are all based on the same general considerations. It should be noticed that although we can change a small E.M.F. into a large one, we cannot create energy. The change is always accompanied by the conversion of a large current into a small one, and the product of E.M.F. and current, which gives the energy, is never increased.

The Induction Coil

In the *induction coil*, which is a special type of transformer, an arrangement is provided for suppressing the E.M.F. created in the secondary when the primary current is started, and increasing that created when the primary current is stopped. Instead, therefore, of having an *alternating* E.M.F. (as an E.M.F. which repeatedly changes direction is called) we have an *intermittent direct* E.M.F. in the secondary, and a moderate E.M.F. applied to the primary can in this way be made to produce a secondary E.M.F. of tens of thousands of volts.

Alternating Currents

It is, however, not always necessary to keep the current constantly in one direction. For certain purposes an alternating current has decided advantages. As a general principle we may note that any effect which depends on the *square* of the current strength is produced with both direct and alternating currents, whereas an effect directly proportional to the current strength (or to an odd power of it) is produced only by a direct current. For example, the heating or lighting effect of a current C amperes in a wire of resistance R ohms is C^2R joules per second, and is produced

no matter how the current may change direction provided that the square of its strength remains equivalent, on the average, to that of a direct current of C amperes. Electrolytic deposition, on the other hand, is simply proportional to the current strength, and a symmetrically alternating current will give no effect, the ions being simply pushed to and fro in the solution with no resultant progress. Mathematically this is associated with the fact that C^2 is positive whether C be positive or negative.

If we wish to make use of electromagnetic induction to produce E.M.F.'s, we are bound to produce alternating ones because the induced E.M.F. exists only while the magnetic flux varies; and since we cannot indefinitely increase a magnetic flux or reduce it below zero, the only way of making it vary for a sufficient length of time is to make it alternately increase and decrease. The resulting current is, of course, alternating too, and unless it is necessary to *rectify* it (*i.e.* convert it into a uni-directional, or *direct* current) we use it as it is.

Mean and Virtual Values of Alternating Current

Clearly, however, an alternating current does not merely oscillate in direction; it oscillates in strength also, and it is important to know how best to estimate the *average* strength appropriate to a given problem. This will depend on the character of the alternation. Thus, if we plot a diagram showing how the current strength changes with time, we obtain a wavy curve, and, just as in the case of sound waves (see I, 211), the form of the wave may vary. In practice, we always seek to obtain a sine wave, such as that shown in I, 199, and reproduced in Fig. 128, but here the ordinates represent current strength and the abscissae time. The equation to the curve is

$$C = C_m \sin 2\pi \frac{t}{T} \quad . \quad . \quad . \quad . \quad (12{\cdot}26)$$

where C is the current strength at any instant t, C_m is the

maximum strength (the *amplitude* of the curve) and T is the period of alternation of the primary (and secondary) current. A similar equation applies, of course, to the E.M.F. as well as to the current.

It will help to fix our ideas if we consider a particular method of obtaining an alternating E.M.F. Suppose the coil in Fig. 124 has no battery in its circuit, but is made to rotate uniformly about the axis XY in the constant magnetic field A. When its plane is parallel to A (the position shown in the figure), it will embrace no lines of force, and as it passes out of this position the rate at which the flux through

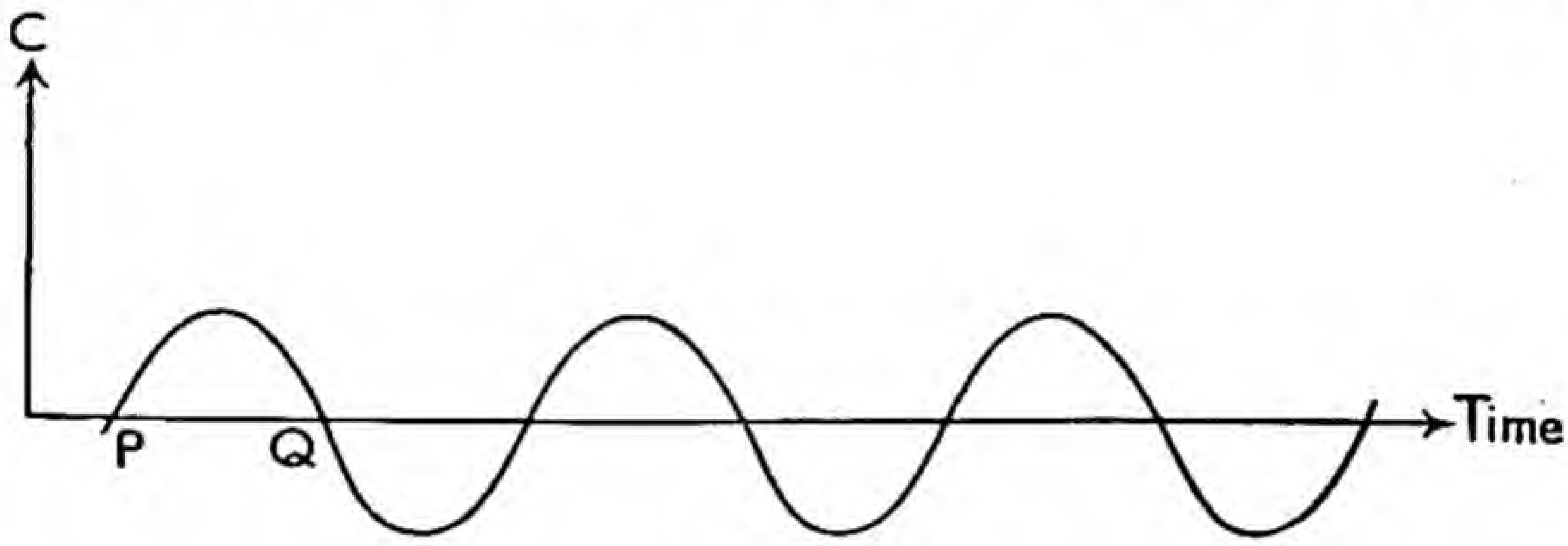


FIG. 128
Sine Curve representing variations in strength of Alternating Current

it changes will be a maximum. The E.M.F. induced is then E_m. After the coil has moved through 90° it takes in as many lines of force as it possibly can, and the instantaneous *rate of change* of this number is zero. After movement through a further 90°, the rate of change of flux is again a maximum, but Fleming's right-hand rule shows that the resulting E.M.F. is in the opposite direction to the previous maximum ; and so on throughout the revolution. The E.M.F., in fact, is represented by Fig. 128. Such a coil, when the magnetic field is that of the Earth, is known as an *earth inductor*. When it is placed between the poles of a permanent magnet and contains many turns of wire wound on an iron core, it is called an *armature*.

In equation (12·26), T is the period of rotation of the

coil; if the coil rotates n times a second, $T = \frac{1}{n}$. The angular velocity is then $2\pi n$ radians a second $= \omega$, say, so that equation (12·26) may be written

$$C = C_m \sin \omega t \quad . \quad . \quad . \quad . \quad (12\cdot27)$$

The average value of this current, if direction be taken into account, is, of course, zero, because it is as frequently, and with the same strength, in one direction as in the opposite one. If we take half a cycle, however—from P to Q in Fig. 128—the average strength $\overline{C}$ is given, by the principles of the differential calculus, as

$$\overline{C} = \frac{1}{\pi}\int_0^{\pi} C_m \sin \omega t \cdot d(\omega t) = \frac{2}{\pi} C_m \quad . \quad . \quad (12\cdot28)$$

This, however, is of little practical importance because, as we have just said, its effect is immediately neutralized by that of the same average in the opposite direction. The proper average for physical effects arising from alternating currents is that of C^2, and this is given by

$$\overline{C^2} = \frac{1}{\pi}\int_0^{\pi} C_m^2 \sin^2 \omega t \cdot d(\omega t) = \frac{C_m^2}{2} \quad . \quad . \quad (12\cdot29)$$

$\frac{C_m}{\sqrt{2}}$, the square root of this, is known as the *virtual* current. This means, for example, that an alternating current whose maximum value is C_m, has the same heating effect as a direct current whose steady value is $\frac{C_m}{\sqrt{2}}$.

The subject of alternating currents is a large and important one. For its further consideration the student is referred to books on electrical engineering, in which it plays a very

prominent part. The application of Ohm's law to alternating currents, for example, is complicated by the occurrence of self-induction in the circuit—a difficulty which does not arise when the current is steady. It may be shown that, instead of the simple relation, $C = \frac{E}{R}$, we have to write

$$\bar{C} = \frac{\bar{E}}{\sqrt{R^2 + \omega^2 L^2}} \quad . \quad . \quad . \quad . \quad . \quad (12{\cdot}30)$$

where $\bar{C}$ and $\bar{E}$ are the virtual values of the current and E.M.F. Instead of the resistance R, therefore, we have a larger quantity, $\sqrt{R^2 + \omega^2 L^2}$, known as the *impedance*.

EXERCISES

1. Find an expression for the magnetic field at the centre of a circular coil carrying a current. If the coil has a radius of 15 cm. and has 10 turns of wire, what current must flow through it in order that, when it is set in the Earth's magnetic meridian, a magnetic needle at its centre shall be deflected through 30° ? The horizontal component of the Earth's field may be taken as 0·18 C.G.S. units.

2. Describe the chief electrical units in the electromagnetic, electrostatic, and practical systems, and state the relations between the units in the three systems. Why is it necessary to use more than one system ?

3. What is the pole strength of a bar magnet 15 cm. long which has the same magnetic moment as a circular coil of 50 cm. diameter and 20 turns of wire carrying a current of 5 amperes ?

4. Prove that the magnetic field inside, and away from the ends of, a long, thin solenoid having n turns of wire per cm., and carrying a current of C e.m. units, is $4\pi nC$ C.G.S. units.

5. What is electromagnetic induction? Describe the experimental facts on which it is based, and give rules for finding the direction of the induced current or motion, as the case may be.

6. Explain the production of eddy currents.

7. State Lenz's law, and apply it to the determination of the directions of the E.M.F.'s induced in a circuit in which a current is suddenly started and stopped.

8. Describe the principle of the transformer. How would you construct such an instrument to obtain (*a*) a large E.M.F. from a small one; and (*b*) a small E.M.F. from a large one?

9. What is an alternating current? Describe a method of producing such a current.

CHAPTER XIII

ELECTRICAL INSTRUMENTS

ELECTRICAL instruments are so numerous and various that it is impossible here to do more than describe the principles and general forms of the chief ones. It often happens that an instrument is permanently enclosed in an opaque casing, and only a couple of terminals are exposed to sight, so that it is impossible to tell by mere inspection what kind of instrument it is. For that reason it is specially important to understand the principles of the instrument, so as to be able to imagine what is taking place inside.

Measurement of Potential Difference

In principle the simplest form of instrument for measuring P.D.'s is the *absolute electrometer*, in which two horizontal metal plates, placed one below the other, are respectively raised to the potentials whose difference is to be measured. In this connection it must be remembered that the potential over a conductor is everywhere the same, so that if the plates are connected by conducting wires to other bodies, they will respectively assume the potentials of those bodies. As a result the plates attract one another with a force proportional to the P.D. between the bodies, for the P.D. is proportional to the work which would be done if the plates were brought together, *i.e.* to the product of the force and the distance between the plates. Hence, if the upper plate forms one "pan" of a balance which is in equilibrium when both plates are at the same potential, the weight which must be added to the other pan to restore balance when the plates are given the potentials to be compared gives a measure of the P.D.

The Quadrant Electrometer : An instrument more often used for this purpose, however, is the *quadrant electrometer*, which consists of four quadrants cut from a cylindrical metal box and detached from one another in the manner shown in Fig. 129. The components of each pair of opposite quadrants are connected by a conducting wire, so that they necessarily come to the same potential. In the centre of the system, but not touching the quadrants, a horizontal metallic "needle" is suspended by a conducting wire. This has the shape shown by the dotted outline in Fig. 129, and its normal position is such that its axis lies along one of the lines of separation between neighbouring quadrants.

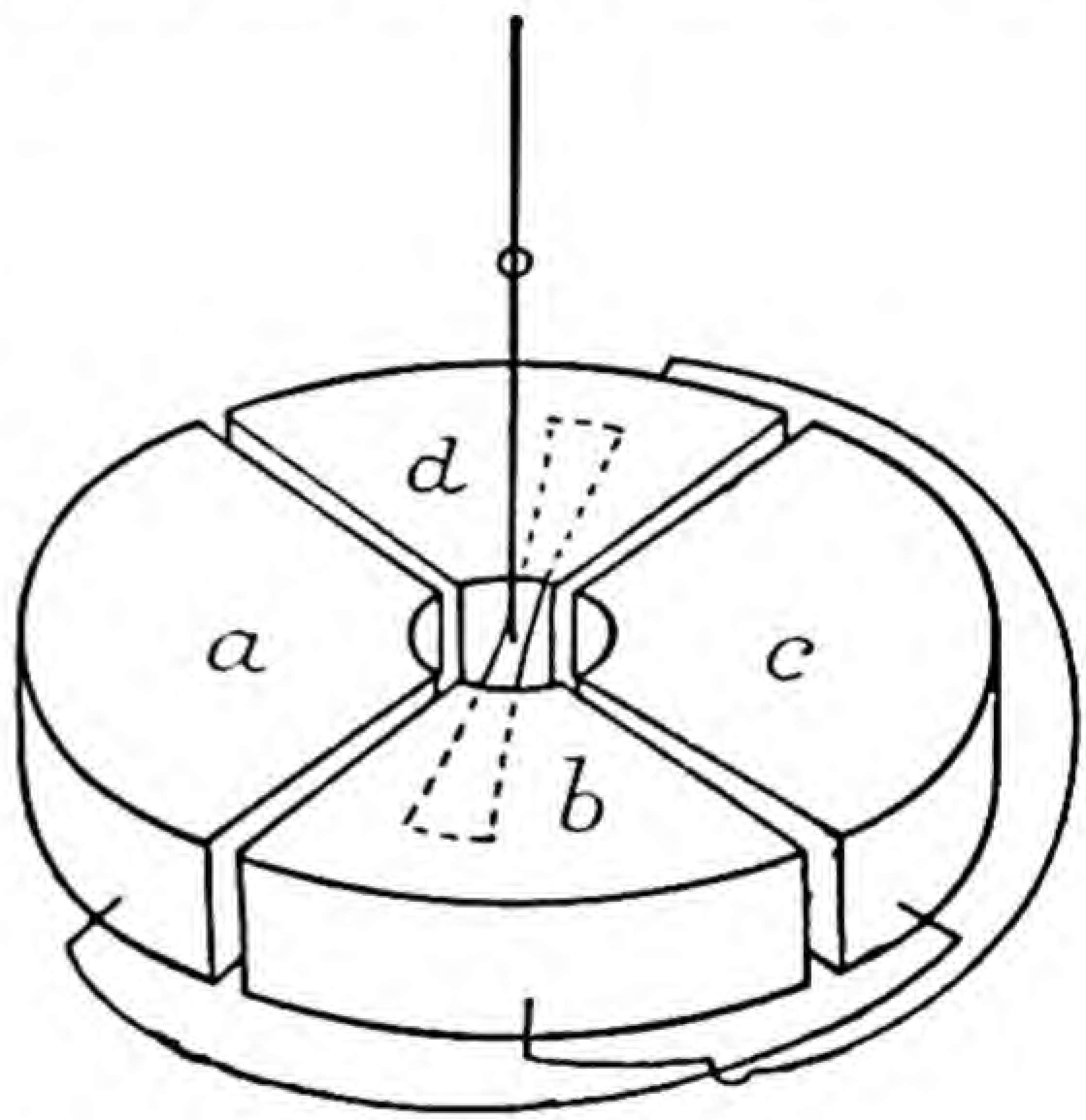


FIG. 129
The Quadrant Electrometer

When the instrument is to be used, the needle is charged (positively, say) to a high potential, and the bodies whose potentials are to be compared are connected respectively to the pair *a*, *c*, and the pair *b*, *d* of quadrants. There will then be a force between the needle and each pair of quadrants—a force of attraction if they are dissimilarly and of repulsion if they are similarly charged—and the magnitude of this force will in each case depend on the P.D. between the quadrants and the needle. The needle will therefore move towards the pair of quadrants which attracts it most strongly, or away from the pair that repels it most strongly. This motion, however, creates a twist in the suspending wire which is opposed by the rigidity of that wire (see I, 54), and the needle finally comes to rest in a position such that the moment of the restoring couple due to the twist is equal to that of the

couple created by the electrical force. If the needle is then deflected through an angle θ from its normal position, it can be shown that

$$\theta = \alpha V(V_2 - V_1) \quad . \quad . \quad . \quad . \quad (13{\cdot}1)$$

where V is the potential of the needle, V_2 and V_1 are the potentials to be compared, and α is a constant. αV may be

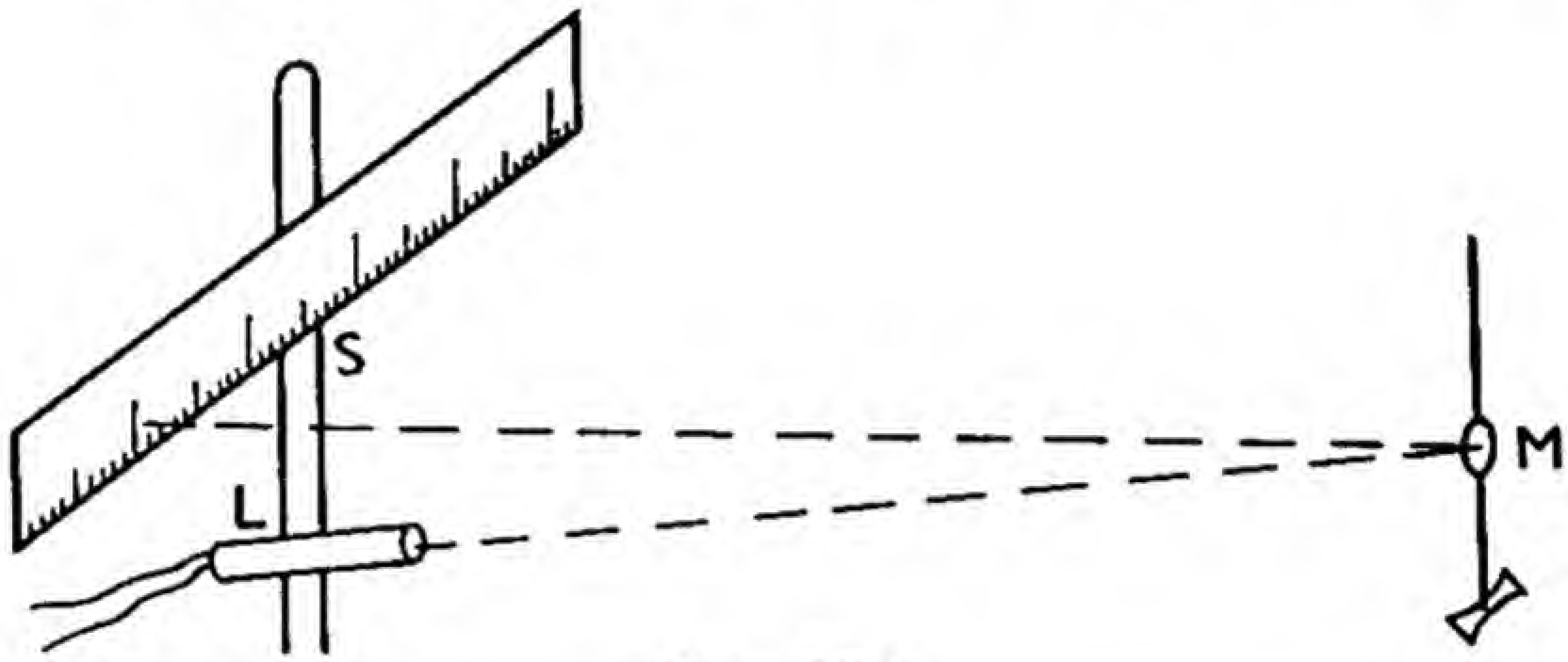


FIG. 130
Lamp and Scale Method of observing Deflections

determined once for all by observing the deflection when the quadrants are joined to the terminals of a battery of known E.M.F., and all the data are then available for measuring $V_2 - V_1$ in terms of θ.

The "Lamp and Scale" Method: The method of observing θ is a very general one, used almost invariably when it is necessary to measure the angle of deflection of a suspended system ; it is known as the " lamp and scale " method. A small plane mirror M is rigidly attached to the suspending wire, and a lamp L, about a metre away, throws a parallel beam of light on the mirror, from which it is reflected to a finely graduated scale S, just above the lamp (see Fig. 130). The system is arranged so that, when the needle is undeflected, the light reflected from the mirror is perpendicular to the scale and falls at its zero point. When deflection through an angle θ takes place, the reflected beam rotates through

an angle 2θ (see p. 45), and if the scale reading is now s, we have

$$\frac{s}{SM} = \tan 2\theta \quad . \quad . \quad . \quad . \quad . \quad (13\cdot2)$$

where SM is the perpendicular distance from the mirror to the scale. In this way small deflections can be accurately observed.

The Electrostatic Voltmeter: When the P.D. to be measured is large, the instrument is somewhat modified. The needle and quadrants are turned into a vertical plane, and the needle is connected to one pair of them; it therefore acquires their potential instead of having a high independent potential of its own. The principle is the same as before, but there are some further differences of detail. The instrument in this form is known as an *electrostatic voltmeter.*

The Potentiometer: A simple method of measuring small differences of potential is afforded by the *potentiometer.*

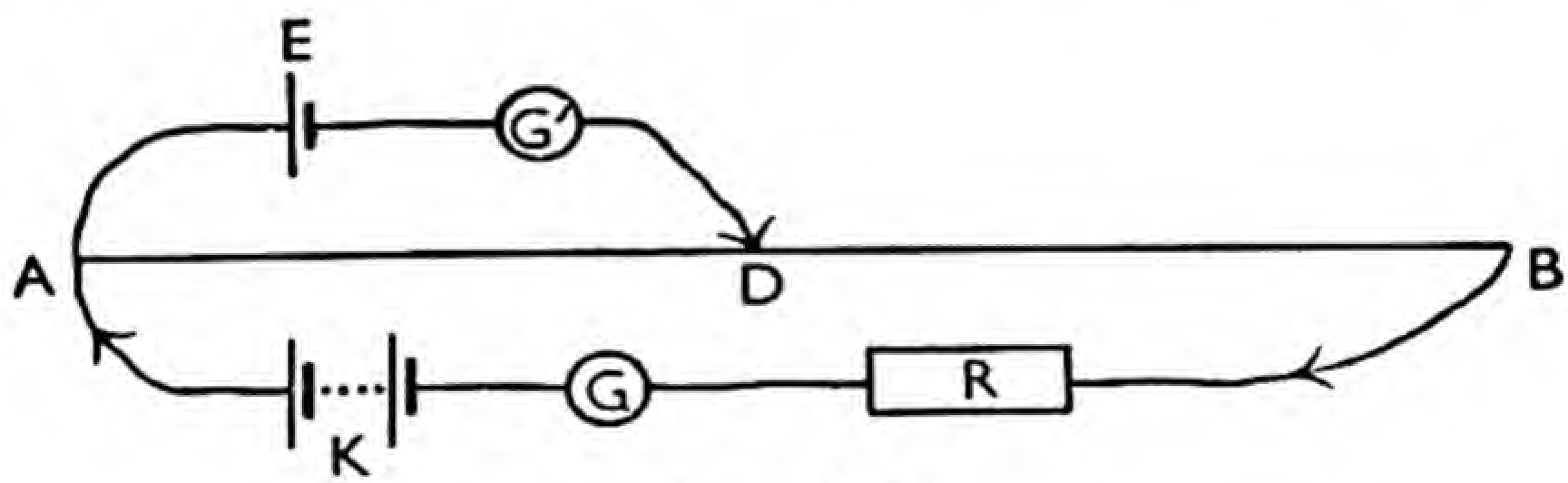


FIG. 131
The Potentiometer

AB (Fig. 131) is a uniform wire forming part of a circuit containing a battery K, a current-measuring instrument (a "galvanometer") G, and a resistance R, the function of which is to prevent the current in the circuit from becoming excessive. Suppose A is joined to the positive terminal of the battery. If the total resistance of AB is r, and the current indicated by G is C, then the P.D. between A and B is Cr. Furthermore, since AB is uniform, the potential falls uniformly

along it, so that the potential at any point D on AB is $Cr \times \frac{AD}{AB}$ below that at A.

Now suppose the P.D. to be measured is that between the terminals of a cell E, and suppose that this P.D. is less than Cr. Then, if the positive terminal of E be connected to A, and the negative terminal attached, through a galvanometer G′, to a loose lead, the free end of which may be moved along AB, a point may be found on AB such that when the lead touches it, thus completing the circuit through G′, that galvanometer nevertheless indicates no current. This point (let it be D) is that which differs in potential from A by the P.D. between the terminals of the cell ; for, in that case, the end of the lead had the same potential as D before they touched, so that contact makes no difference. The P.D. between the terminals is then the E.M.F. of the cell, which is therefore equal to $Cr \times \frac{AD}{AB}$.

Measurement of Electric Current

An instrument for measuring an electric current is called a *galvanometer*. Portable instruments, used for comparatively rough measurements, are known as *ammeters*.

The Tangent Galvanometer : This instrument consists of a small magnetic needle suspended at the centre of a circular coil, standing in a vertical plane, through which a current can be made to flow. If the coil is placed in the Earth's magnetic meridian, the needle will lie in the plane of the coil when no current is flowing through the coil. When a current C flows, however, a magnetic field perpendicular to the Earth's horizontal component is created at the centre of the coil, of magnitude

$$F = \frac{2\pi nC}{R} \quad . \quad . \quad . \quad . \quad . \quad (13{\cdot}3)$$

(from (12·8), with C in e.m. units so that $k = 1$), where R is the radius and n the number of turns in the coil. The needle

is therefore deflected, and comes to rest when the moment of the deflecting force F is equal to that of the horizontal component H of the Earth's field, which tends to keep the needle in its original position. We have here a situation similar to that discussed on p. 192, and, as before, we have for equilibrium,

$$\left.\begin{aligned} F \cos \theta &= H \sin \theta \\ \text{or } F &= H \tan \theta \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (13\cdot4)$$

where θ is the angle of deflection.

$$\left.\begin{aligned} \text{Hence } \frac{2\pi nC}{R} &= H \tan \theta \\ \text{or } C &= K \tan \theta \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (13\cdot5)$$

$$\text{where } K = \frac{RH}{2\pi n} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13\cdot6)$$

This quantity is constant for a given coil provided that the Earth's field does not vary. It does vary somewhat, however, so that, for accurate work, either the instrument should be calibrated by passing a known current through it, and so determining K from the observed value of θ, or else H should be determined at the time and place of use. K is, nevertheless, often called the *constant* of the galvanometer.

The instrument is most accurate when the radius of the coil is large and the needle is small, so that F is practically constant over the whole range of swing. The suspending wire should be long and thin, otherwise the restoring couple resulting from torsion in it (which, in the quadrant electrometer, was the only restoring couple, but here is comparatively so small as to be negligible) will have to be added to that due to the Earth's field. The deflection is read by the lamp and scale method, or sometimes by a long, light pointer fixed at right angles to the magnet and moving over a horizontal scale of degrees.

The Suspended Coil Galvanometer: This is the commonest form of galvanometer for ordinary laboratory work. It consists of a rectangular coil containing a large number of

turns of fine conducting wire (see Fig. 132, which, however, shows a few turns only, too widely separated) suspended between the poles of a permanent magnet. The two ends of the coil are attached, the upper one to the suspending wire and this to an external terminal, and the lower to a spring from which a lead passes to another terminal. The instrument is inserted through these terminals into the circuit in which flows the current to be measured, and the current thereupon

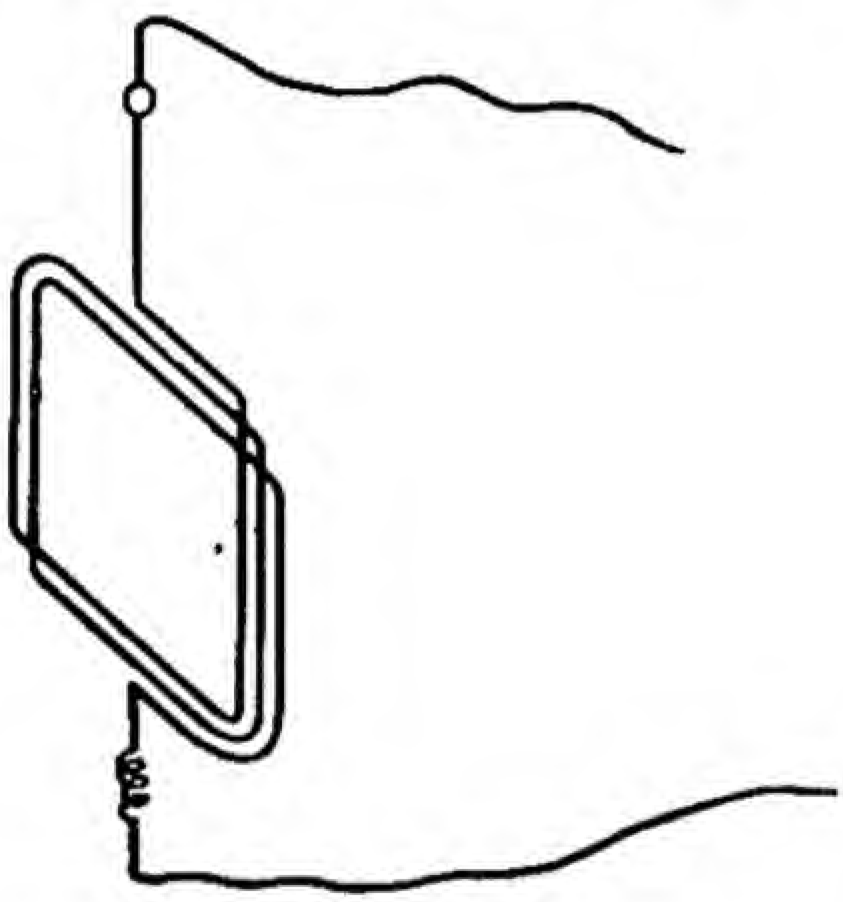


Fig. 132
Coil of Suspended Coil Galvanometer

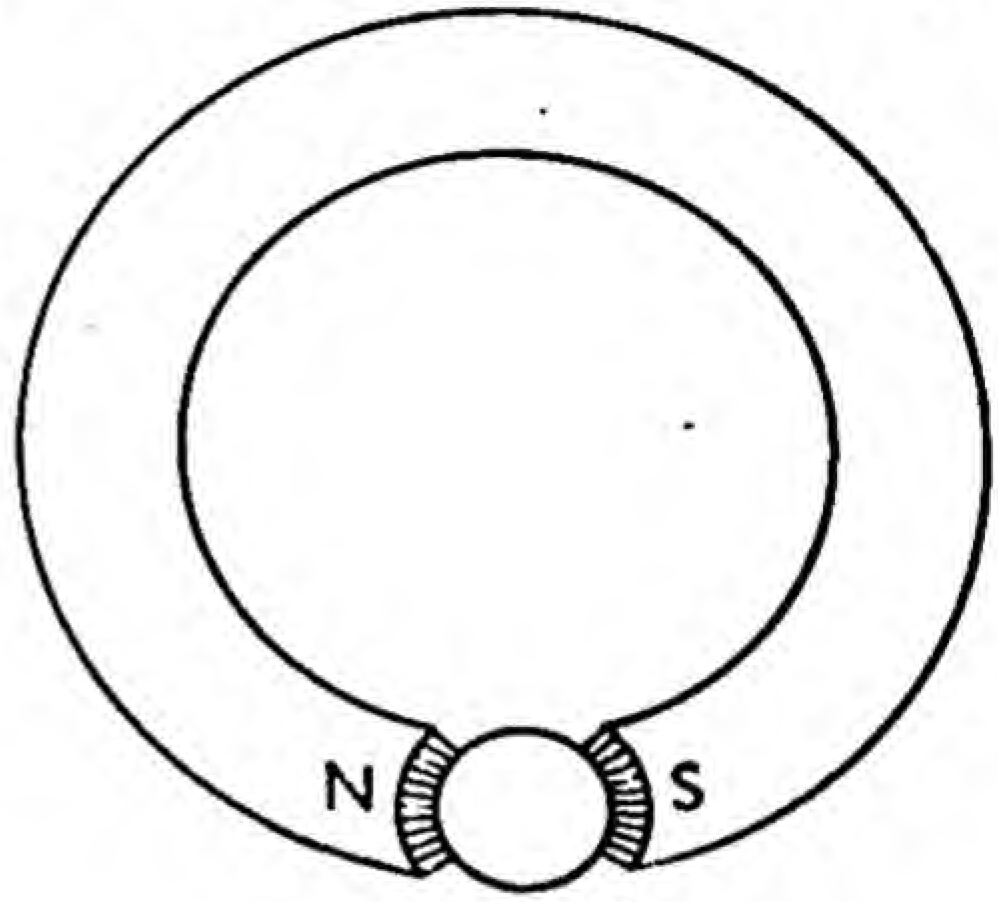


Fig. 133
Magnet of Suspended Coil Galvanometer

flows through the coil. The instrument is set so that, when no current is flowing, the coil is parallel to the field of the magnet. When a current flows, however, the coil tends to turn so as to embrace as many lines of magnetic force as possible (p. 216), but this tendency is opposed by the torsion of the suspending wire, so that it comes to rest when the moments of the opposing couples are equal. The greater the current the greater is the deflecting couple, and therefore the greater the angle of deflection when the system comes to equilibrium. The deflection is thus a measure of the current.

In order that the magnetic field shall be as uniform as possible, the magnet is in the form of a short vertical cylinder with the central portion hollowed out (Fig. 133 shows a

section by a horizontal plane) and slits of circular section cut for the coil to move in. The lines of force are then normal to the cut surfaces, and form a uniform field as long as the deflection of the coil is not too great.

The relation between current and deflection may be determined as follows. We have seen (p. 213) that a coil carrying a current acts as a magnet with a moment equal to the product of the current and the area of the coil. If the area of each turn is A, and there are n turns, the moment of the equivalent magnet is therefore nAC. (We have actually proved that the coil has this moment only in so far as its action at distant points is concerned, but the result does in fact hold good also for the application here made of it.) Owing to the shape of the pole-pieces of the magnet, the coil cuts the lines of force normally in all its positions, so that the turning force due to the magnet is independent of the deflection (so long as that is not excessive) and is proportional to the moment of the equivalent magnet, *i.e.* to the current C. The torsional restoring force, on the other hand, increases with the angle of deflection and is proportional to it (see I, 55). Hence a deflection α will be reached for which the moment of the magnetic force (kC, say, where k is a constant) is equal to $T\alpha$, the moment of the restoring force—T being the moment of the restoring couple for unit deflection. We have then

$$\left.\begin{aligned} kC &= T\alpha \\ i.e. \quad C &\propto \alpha \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (13{\cdot}7)$$

For small deflections α is proportional to $\tan \alpha$, which is proportional to the scale reading when the lamp-and-scale method is used. Hence the current is proportional to the scale reading.

The String Galvanometer : A more delicate instrument, used only for certain special work, is the string galvanometer, which is similar to the suspended coil galvanometer except that instead of a coil there is a single string of fine quartz silvered to make it conducting. The string is held taut, and,

when a current passes through it, it moves in the magnetic field according to Fleming's left-hand rule, the force acting on it being proportional to the current. The amount of movement is observed through small holes in the pole-pieces of the magnet.

The Ballistic Galvanometer: It is sometimes desired to measure, not the steady current flowing in a circuit (the quantity of electricity passing each point of the circuit per second) but the total amount of electricity which passes in a sudden discharge, as when the two plates of a charged condenser are joined, irrespective of the (very short) duration of the discharge. In such a case a suspended coil galvanometer can be used, in which the suspended system is constructed so as to be slow in responding to the passage of electricity, so that the discharge is complete before the coil has moved appreciably from its normal position. In such a case the impulse it receives results in its being deflected through a certain angle θ and then oscillating slowly to and fro before finally returning to rest in its initial position. The extent of the first swing (or "throw," as it is usually called), after a correction, determined from separate investigations, has been applied for damping, is a measure of the impulse given it, which in turn is a measure of the quantity of electricity which has passed. It can be shown that this quantity is proportional to $\sin \frac{\theta}{2}$. A galvanometer used in this way is known as a *ballistic* galvanometer.

Electrical Measurement of Radiation

The most sensitive methods of measuring heat or light radiation are electrical in principle, depending on electrical effects of the heat produced when the radiation is absorbed by a solid body. The thermopile has already been described (p. 159). We shall now describe two other instruments—the *radio-micrometer* (the most sensitive of all) and the ***bolometer***.

The Radio-micrometer : This instrument, invented by Sir C. V. Boys, is a combination of a thermocouple and a moving coil galvanometer. The coil of such a galvanometer consists here of one turn of copper wire whose lower ends are attached to small bars of antimony and bismuth connected through a blackened copper disc (Fig. 134). The radiation is allowed to fall on this disc, and heats the junction of the antimony and bismuth so that a thermo-electric current is produced, and this tends to turn the coil in the field of the magnet. The system is suspended by a quartz fibre (which has the property of returning almost immediately to its original position when the deflecting force is removed), and the deflections are read by the lamp-and-scale method.

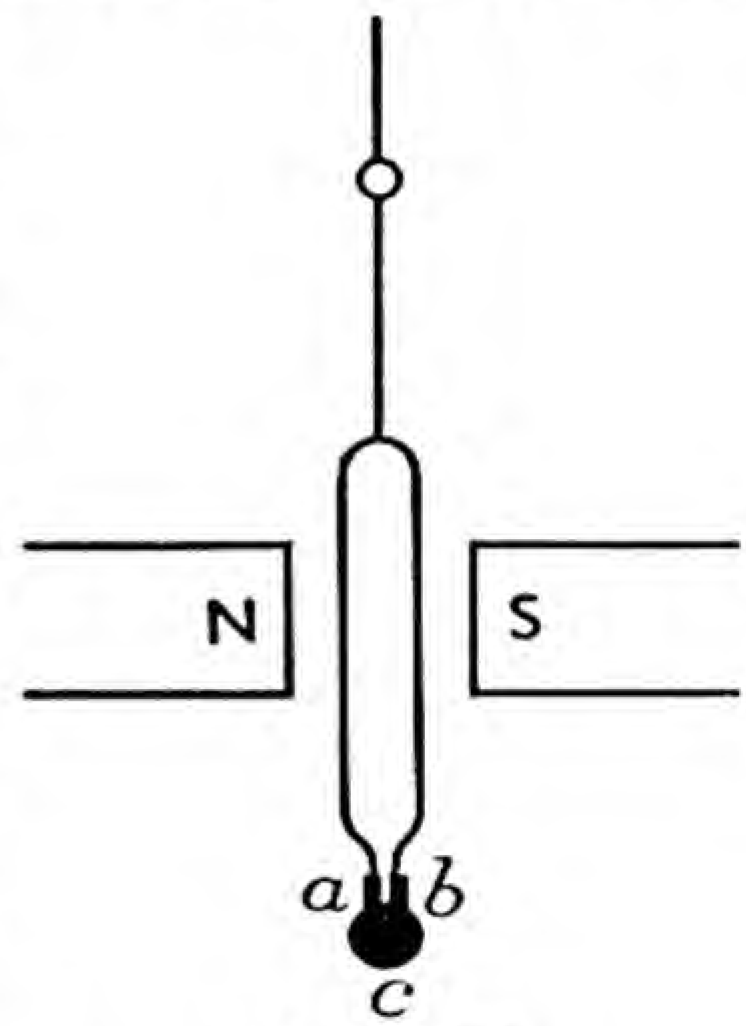


FIG. 134
The Radio-micrometer
(*a*) Antimony
(*b*) Bismuth
(*c*) Blackened Copper Disc

The Bolometer : This instrument, now less used than formerly, consists of a thin strip of blackened platinum placed in one of the gaps of a Wheatstone's bridge. If the bridge is balanced at uniform temperature, the balance will be disturbed if the platinum is heated by the reception of radiation, since the resistance of the metal increases with temperature (pp. 170–71). The increase of resistance is approximately proportional to the increase of temperature, which, in its turn, is proportional to the amount of radiation received when that is not too large.

The Post Office Resistance Box

This instrument gives, in a convenient form, a means of inserting in a circuit a resistance of any integral number of ohms up to a maximum depending on its construction. Its principle may be understood from Fig. 135, which shows a

portion in diagrammatic form. A number of thick brass blocks, A, B, C, D, . . . of negligible resistance, are separated by spaces, any or all of which can be closed by thick metal plugs such as T. Neighbouring blocks are connected underneath by coils, *a*, *b*, *c*, . . . each of which has a known resistance of an integral number of ohms. The current enters the instrument at P, and leaves it at the last block by a wire such as that shown in Fig. 135 as proceeding from D. If all the plugs are in the spaces between the blocks, the resistance of the whole box is effectively zero, but if any plug is removed, the current must go through the coil below, and the resistance

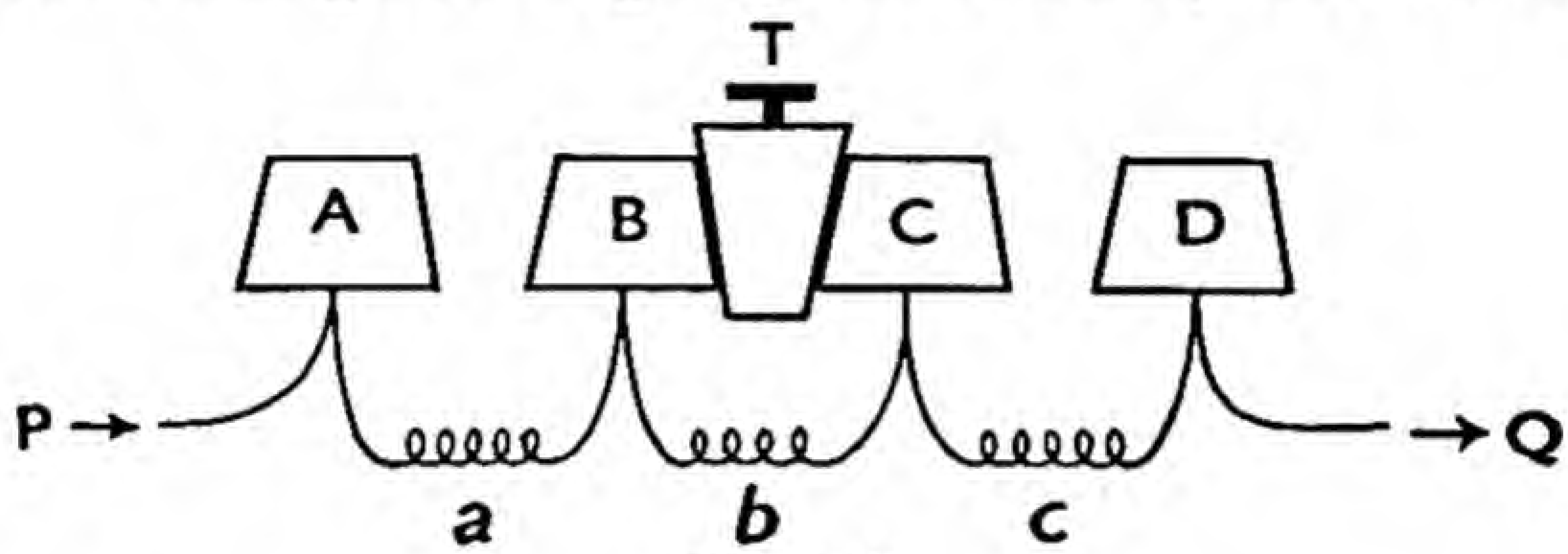


FIG. 135
Portion of Post Office Resistance Box
(As shown, resistance in circuit $= a + c$)

of that coil is thereupon inserted in the circuit. By having coils of 1, 2, 2, 5, 10, 20, 20, 50, 100, 200, 200, 500, and 1,000 ohms in the box, it is possible to obtain a total resistance of any integral number of ohms up to 2,110.

The Carbon Rheostat

It is sometimes desired to vary the resistance in a circuit continuously, *i.e.* by amounts of *any* magnitude below a certain limit, and not merely in multiples of a unit. For this purpose a *carbon rheostat* is employed, consisting of a number of thin flat rectangular plates of carbon placed side by side, with a metal plate at each end carrying terminals for including the instrument in the circuit (Fig. 136). By means of a screw a variable pressure *P* may be applied to the system so as to bring the plates into closer or less close contact, as may be

required. Contact between the rough surfaces of the plates is not uniform all over, but occurs only at certain points, and as the pressure is increased, the total area of contact is enlarged. The resistance is thereby decreased, being inversely propor-

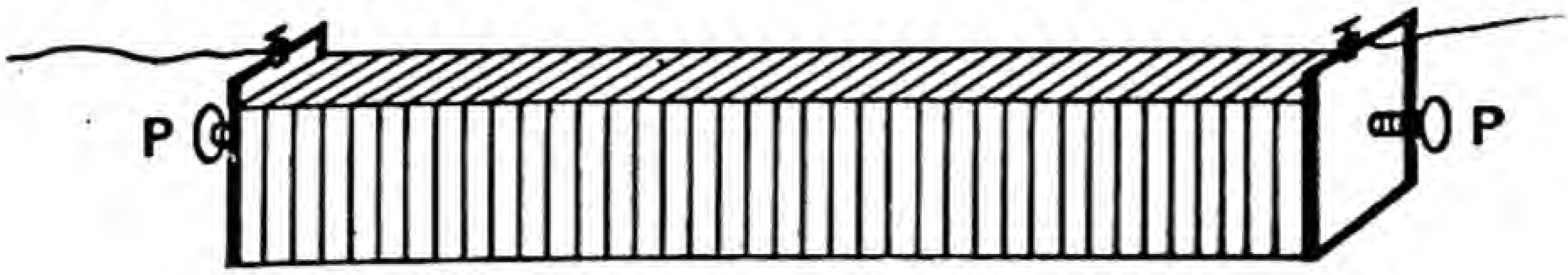


FIG. 136
Carbon Rheostat

tional to the cross-section of the conductor (p. 169). By adjusting the pressure to the proper value, the desired resistance can therefore be obtained. It is not, of course, measured, but an indication that the desired conditions have been reached is usually given by a galvanometer which shows that the required current is flowing.

The Telephone

The telephone system between two widely separated points, A and B, consists of two similar circuits, one at A and

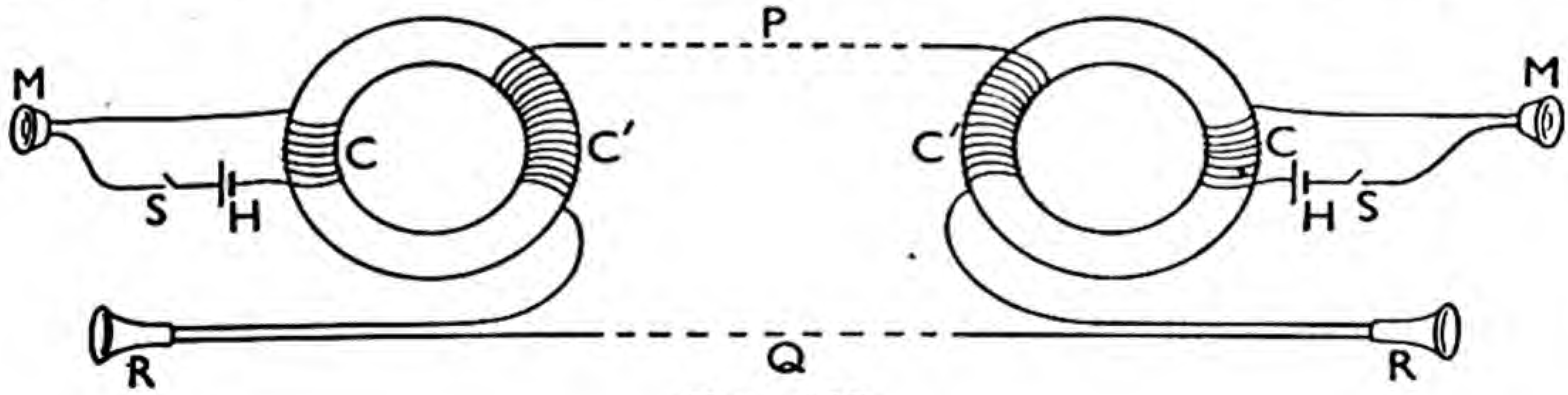


FIG. 137
Telephone Circuit

C, C′ Primary and Secondary Coils
H Batteries
S Switches
M Microphones
R Receivers

the other at B, each of which forms the primary circuit of a transformer. The secondary circuit (PQ in Fig. 137) is common to both transformers. Each primary circuit con-

tains a battery H, a switch S, a microphone M, into which the user speaks, and a coil C, while the secondary circuit includes the two coils C′, having many more turns than the primary coils C, and at each station a receiver R, in which the listener hears the message.

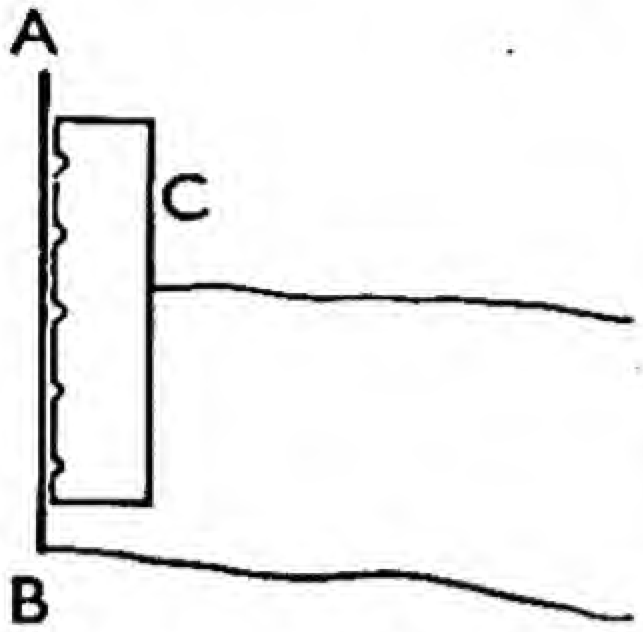


FIG. 138
The Microphone

The Microphone: The microphone is an instrument designed to make the current in the primary circuit vary with variation in the sound spoken into it. Its essential parts, shown in Fig. 138, are a very thin carbon plate, AB, placed very close to a block of carbon C, at intervals along which small cavities are made and loosely filled with coarsely powdered carbon. The contact thus established between the plate and the block through the carbon grains can be varied considerably if the distance between AB and C is slightly altered. The sound of the speaker's voice falls on the plate, which responds to the alternations of pressure in the sound wave, and so varies, or *modulates*, the resistance and consequently the current in the primary telephone circuit. The E.M.F. induced in the secondary, which contains the receiver, is therefore similarly modulated, and the receiver is designed to re-convert these modulations into variations of sound, thus reproducing the speaker's voice.

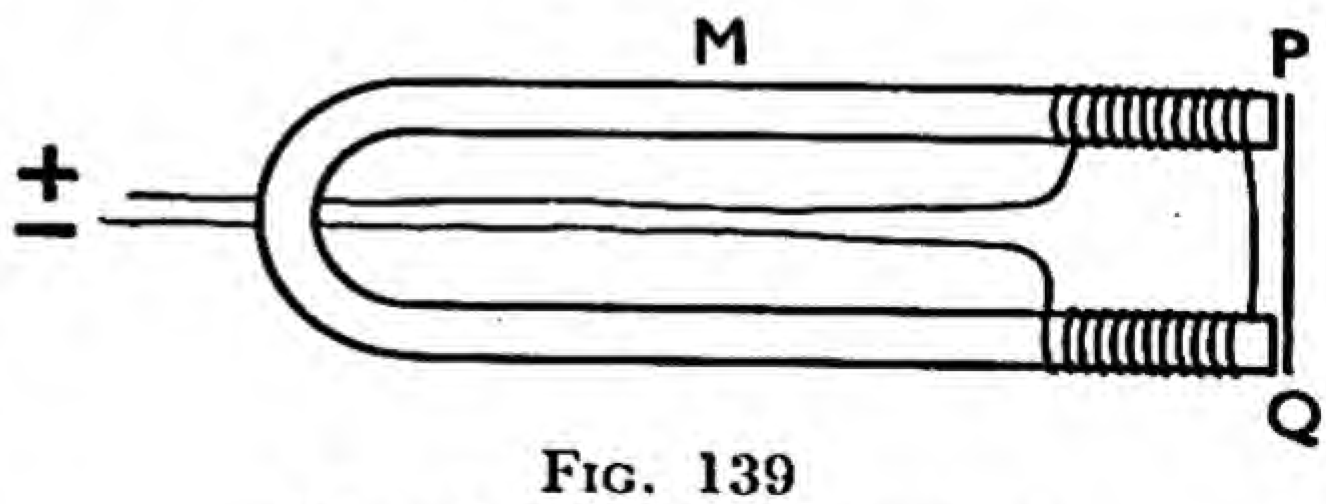


FIG. 139
The Telephone Receiver

The Receiver: The receiver is illustrated in Fig. 139. It consists of a thin soft iron diaphragm, PQ, placed very close to a U-shaped piece of soft iron M, round the ends of which are wound coils forming part of the telephone secondary circuit, and there-

fore carrying the modulated current. As the current varies, the strength of the magnetic poles which it creates in M varies, and therefore also the attraction which they exert on PQ. Hence PQ vibrates with the frequency and variation of amplitude of the modulations, and so sets up sound waves to the right of PQ, which enter the listener's ear.

The principle of the telephone is therefore the conversion of sound into electrical energy and back again, and the details are designed to preserve the frequencies and relative amplitudes of the modulations throughout, so that what is heard resembles what is said.

EXERCISES

1. Describe an instrument for measuring the difference of potential between two conductors.

2. A current passes through a circuit containing a solution of a silver salt and a tangent galvanometer whose constant is 0·2 C.G.S. unit. If the deflection of the galvanometer is 30°, how much silver is deposited in half an hour? The electrochemical equivalent of silver is 0·001118.

3. Describe the suspended coil galvanometer, and explain the lamp-and-scale method of observing its deflections.

4. Give an account of a telephone system capable of operating over great distances.

CHAPTER XIV

ELECTRICAL RADIATIONS AND ATOMIC DISINTEGRATION

Range of Ethereal Waves

IN the first chapter we have described several kinds of radiation, of both wave and particle type. We are now in a position to consider some of them in more detail than before. There is no known definite limit at either end to the possible length of ethereal waves. Very long waves, measured in hundreds of metres, are used in wireless transmission. Shorter waves, from about $\frac{1}{10}$ to $\frac{1}{1000}$ mm., are usually known as infra-red waves, and the radiant heat from terrestrial hot bodies consists mainly of radiations within this range; such radiation has been discussed in I, Chapter XII. Waves between $\frac{8}{10,000}$ and $\frac{4}{10,000}$ mm. constitute visible light, and those shorter than this, down to about 10^{-7} mm., are ultra-violet radiations, which, except that they do not produce the sensation of sight, have, broadly speaking, somewhat similar properties to visible light. When we reach the last-named figure, however, we are in the region of X-rays, which have great penetrating power. Generally speaking, the thickness of a given material through which X-rays will pass decreases with increase of density, dense substances such as lead being penetrable only in small thicknesses. The γ-rays of radium and the ethereal radiations found among cosmic rays are of still smaller wave-length than X-rays, and are the shortest waves yet detected. All this is shown in Fig. 3, and is summarized again here for convenience.

In this chapter we shall briefly consider wireless waves and X-rays, leaving the detailed account of their various

applications to books dealing specially with such subjects. We shall also refer to some aspects of electrons not yet dealt with, and describe some of the results of bombarding atoms by sub-atomic particles.

Electrical Oscillations

If a charged condenser is discharged by connecting the plates by a high or moderately high resistance, the discharge does not take place instantaneously but gradually, the P.D. between the plates steadily decreasing to zero. If, however, the connecting wire has a very low resistance, the discharge is so rapid that it overshoots the mark, so to speak, and the plates become charged in the opposite sense. The condenser immediately discharges again, but several oscillations may occur before it is finally discharged.

The to-and-fro oscillation of charge in these circumstances results in a wave being sent out into the ether having the same frequency, n, as the oscillations, and, of course, a wave-length $\frac{c}{n}$ where c is the velocity with which the ether transmits waves—the velocity of light. In the circumstances of ordinary experiments, n is such as to give very long waves, and this is, in principle, the way in which wireless waves are generated. Many refinements, of course, are introduced in the practical application of the principle, but the origin of the waves is always in the oscillations of charge between bodies of observable size. The shorter, visible waves, as we know, are generated by the movements of electrons in atoms, and the two processes seem to be different in character. Their reconciliation has been one of the great problems of theoretical physics, and although much progress has been made in recent years, it has not yet been completely solved. For practical purposes, however, that is immaterial, and we may consider the generation of light and of electric waves as two separate processes.

The production of electric waves is facilitated by includ-

ing a coil having a self-inductance L, say, in the wire through which the condenser is discharged. If, then, the resistance of the discharging system is small, it can be shown that the frequency of oscillation is equal to $\frac{1}{2\pi\sqrt{LC}}$, where C is the capacity of the condenser, and both L and C are measured in e.m. units. It is this type of discharge that is generally used.

It is not necessary here to deal with the methods by which wireless waves are put to practical use, as these are described in another volume of this series.

X-rays

To explain the production of X-rays we must describe what happens when an electric current is passed through a tube of gas (air, for example), which is gradually being exhausted (Fig. 140). At first the discharge will not pass at all, since,

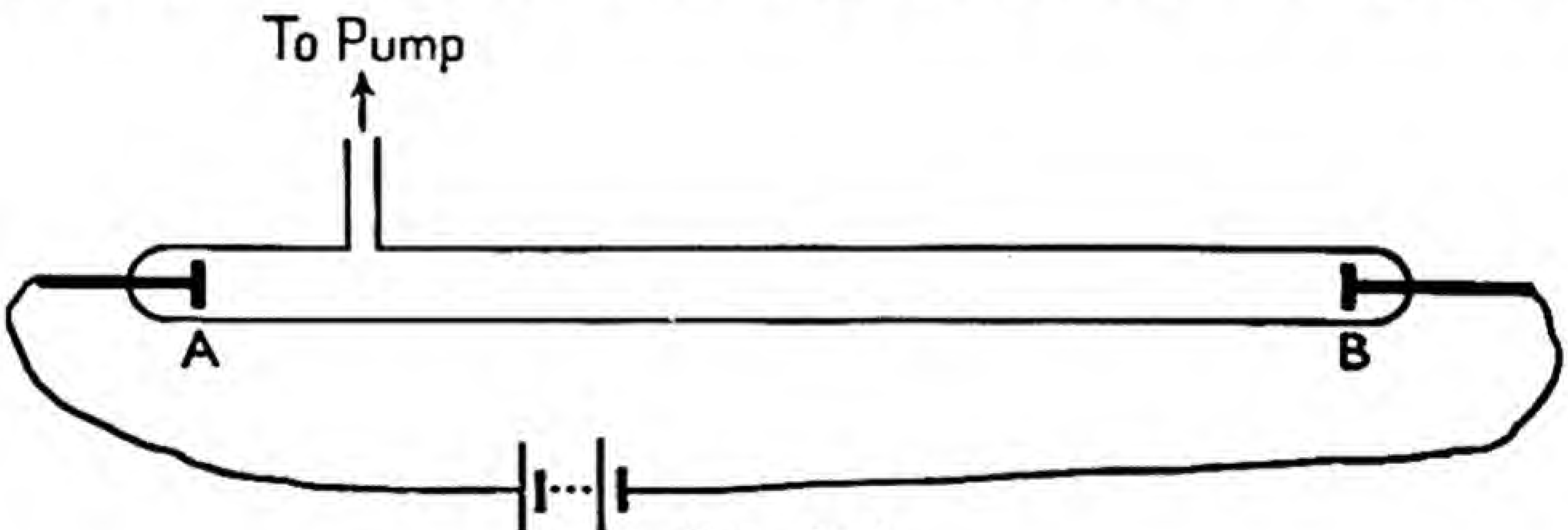


FIG. 140
Gaseous Discharge Tube

at atmospheric pressure, gases are non-conductors of electricity and only allow a momentary current to pass at extremely high P.D.'s such as those which produce lightning flashes and ordinary electric sparks. As the pressure in the tube diminishes, however, the conductivity of the gas increases, and when the pressure has fallen below 10 mm. of mercury a thin spark passes between the positive terminal A (the anode), and the negative terminal B (the cathode). At still lower

pressures this spark broadens until, at about $1 - 3$ mm., it fills the tube with a reddish glow called the *positive column.* As the pressure decreases, this column breaks up into a number of striations, and round the cathode a dark space appears and gradually extends, terminating in a small region of bluish light which separates it from the positive column. The conductivity of the gas now begins to decrease, and at about $\frac{1}{100}$ mm. pressure the discharge almost ceases, the cathode dark space having extended almost throughout the tube. Where it has reached the walls of the tube a greenish fluorescent glow appears on them.

Process of Conduction in Gases : The cause of these phenomena is roughly as follows. Electrons carrying the current along the connecting wire are unable at first, when they enter the tube at the cathode, to acquire sufficient speed to have any effect on the atoms of the gas ; they simply strike those atoms and rebound harmlessly, and the current does not therefore get round the circuit. (The atoms of metallic conductors are far more easily ionized than those of gases, and easily yield electrons which can carry a current.) As the pressure is reduced, however, the atoms of the gas get farther apart, and the electrons have to travel farther before reaching them. During this longer journey they acquire greater speed under the influence of the E.M.F. of the circuit, and when the pressure is sufficiently low they can thus gain sufficient energy to knock electrons out of the atoms of the gas when they at last strike them. These liberated electrons then move under the influence of the E.M.F., and so carry the current round the circuit. Some of them return to the ionized atoms, and in so doing emit light in the manner already described (pp. 22–23). This accounts for the spark which appears in the tube and broadens later into the positive column. As the pressure decreases, the electrons must travel still farther before reaching the atoms of the gas—*i.e.* the cathode dark space extends. When it reaches the

walls of the tube, those walls receive the full impact of the "cathode rays," as these electrons were at first called, and this makes them glow with the greenish colour observed. When at last the pressure in the tube becomes too low for sufficient collisions to occur, the discharge ceases.

Let us now suppose that within the tube, opposite the cathode, there is a metal obstacle (A in Fig. 141), which may be the anode or an independent object. It is then found that when the pressure in the tube is very low, so that the dark space extends from the cathode C to beyond A (the *anti-cathode*, as it is called), the cathode rays on striking A give rise to extremely short ethereal radiations which are

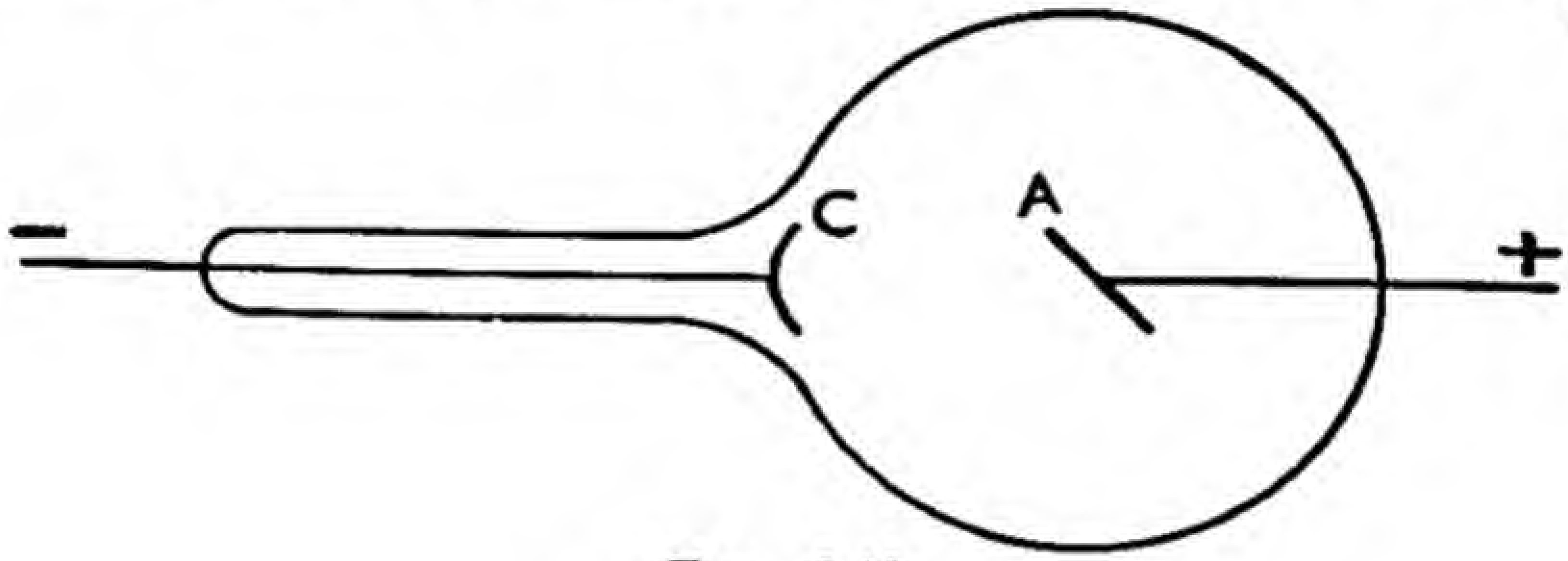


FIG. 141
X-ray Tube

able to escape through the walls of the tube and affect objects (*e.g.* a photographic plate) outside. These are the X-rays. The method of their origination seems to be as follows. An atom of the anti-cathode, as we know, consists of a nucleus surrounded by a number of electrons. If one of the outermost electrons is displaced and returns again, a beam of ordinary visible light is emitted. The cathode rays, however, have sufficient energy to penetrate the outermost shell of electrons and knock out an electron near the nucleus. An electron from one of the outer shells then falls in to take its place, and in so doing emits an X-ray. These rays are therefore emitted in essentially the same way as ordinary light, but by the falling of an electron to a shell within its normal one instead of an original outward displacement followed by

a return to the normal shell. The change of energy is much greater in this case, and the wave emitted is therefore of much higher frequency and smaller wave-length (see equation (1·1).

Hardness of X-rays: The precise wave-lengths of the X-rays vary according to the material of which the anti-cathode is composed. Naturally, the X-ray spectra of the elements differ from one another just as the visible spectra do. The greater penetrating power of X-rays is at least partly due to their greater energy, for, from (1·1), the energy of an ethereal radiation is proportional to its frequency. The higher the frequency of an X-ray, the *harder* it is said to be, and the penetrating power increases with the hardness.

X-rays and Crystal Structure: X-rays have an extremely important field of application in the determination of the structure of crystals. The atoms in a crystal are regularly arranged in layers, forming, in fact, something equivalent to a three-dimensional diffraction grating (see p. 124), with a very small distance between neighbouring opaque spaces, which in this case are atoms. X-rays are diffracted by such a crystal just as visible light is diffracted by a ruled grating, and the paths they take on leaving the crystal allow us to deduce what the structure of the crystal must be. Fig. 91, for example, shows the structure of rock salt deduced in this way.

Photo-electricity

The production of X-rays is an example of ethereal radiations being produced by the impact of electrons on matter. The reverse process also can occur; electrons can be released from matter by the impact of ethereal radiations. In considering this phenomenon, however, it is best to think of the radiation as composed of photons instead of ethereal waves. The simplest example is afforded by the incidence of ordinary light on certain metallic surfaces. When, for instance, red light falls on a caesium surface, or green light on a sodium

surface, electrons leave the metal, liberated from its atoms by the action of the light. This is known as the *photo-electric effect*, or *photo-electricity*.

What happens is the opposite of the emission of light by the inward falling of electrons. A photon, falling on an atom, imparts its energy to one of the outer electrons, and sends it out of the atom. The energy needed to do this varies according to the chemical nature of the atom ; sodium requires more energy than caesium, for instance. If the required energy for a particular atom is E, then the frequency of the incident light must be at least as great as $\frac{E}{h}$, where h is Planck's constant (see Chapter I), and the corresponding photon must have energy at least equal to E. A higher frequency will do (*i.e.* a photon of greater energy) and the higher the frequency above the minimum value the greater will be the kinetic energy with which the electron leaves the metal, but a smaller frequency will have no effect at all. That is why red light, which is effective for caesium, is ineffective for sodium. Some metals need ultra-violet photons to liberate their electrons, and gases need X-rays. Under the influence of X-rays gases become conductors of electricity, for the electrons set free by the photo-ionization are able to carry the current.

An increase of intensity of the incident light without change of frequency does not change the velocity of the liberated electrons, but increases their number. This is the basis of the use of photo-electricity in photometry (see p. 108).

Thermionic Emission

Another way in which electrons can be set free from metals is by simply heating the metals. As the temperature rises, collisions between neighbouring atoms become more violent, and at a certain temperature they have enough energy to break up the atoms and set electrons free. This temperature is of the order of 1,000° C. or higher, so that metals of high

melting-point are required to show the effect. This mode of production of electrons is known as *thermionic emission.*

The Thermionic Valve: The most important application of thermionic emission is found in the *thermionic valve*, on which modern wireless transmission largely depends. This consists of three parts—a *filament* F, a *grid* G, and a *plate* P (Fig. 142, which is purely diagrammatic), all enclosed in an evacuated bulb. When a current passes through the filament, raising it to a high temperature, electrons are released from it, and if the plate has a positive potential, these electrons travel towards it through the grid, thus giving rise to a current in a circuit which is previously made and in which the plate is included. The function of the grid is to control the current in the plate circuit. If it has a positive potential with respect to the filament, it assists in drawing electrons toward the plate, and so increases the current. If, on the other hand, it has a negative potential, it keeps the electrons off. In this way large variations in the current through the plate circuit can be produced by relatively small changes of grid potential. For the application of this instrument to wireless transmission, books on wireless should be consulted.

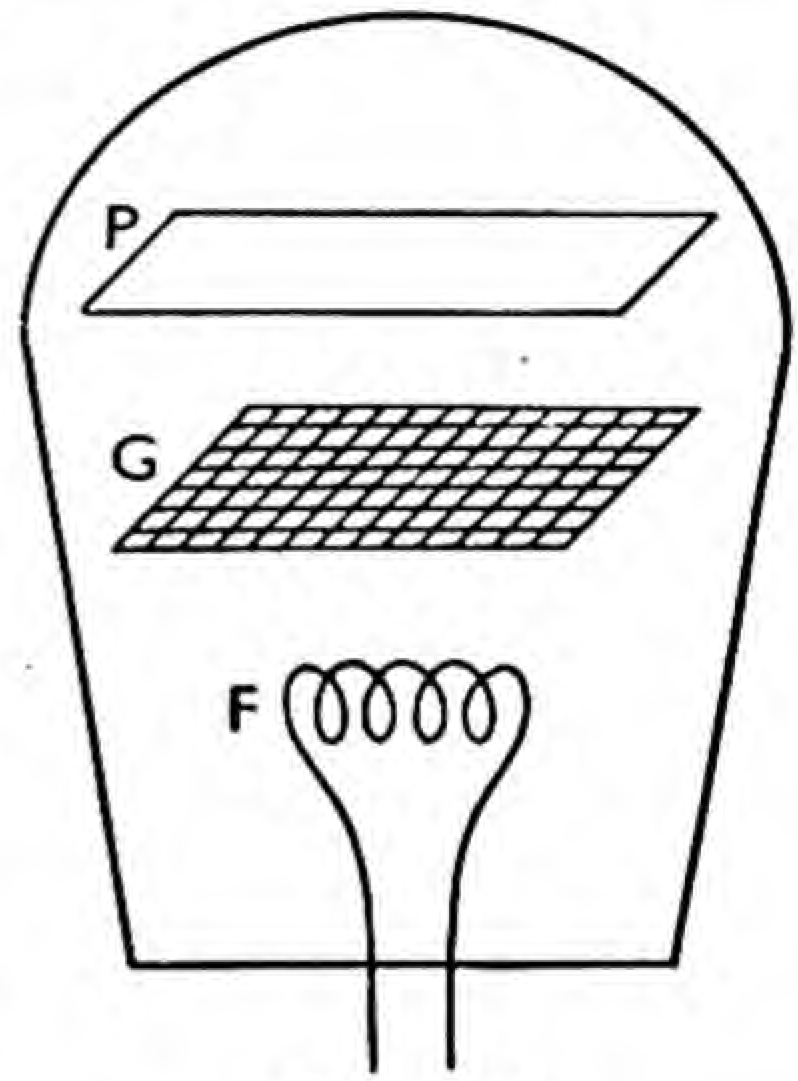


FIG. 142
Diagram of Thermionic Valve
F Filament G Grid
P Plate

Rectification: We may mention the principle of one very important application of the thermionic value—namely, to the rectification of alternating currents; *i.e.* the conversion of an alternating into an intermittent direct current. Let the alternating E.M.F. which produces the current be applied to the grid of a valve so that its potential alternates with

respect to that of the filament. The electrons will then be alternately attracted to and repelled from the grid, and a succession of pulses of electrons will pass through. These electrons will thus set up an intermittent direct current in the plate circuit, giving us the rectified current we require. In practice, the apparatus is more complicated than this, mutual inductance between the circuits also being called into play, but we have said sufficient to make clear the action of the valve. It depends essentially on the fact that the hot filament emits electrons (negative charges) but not positive charges, so that unless electrons reach the plate, nothing reaches it.

Bombardment of Atoms

One of the most remarkable uses to which electrons and other particles have been put is their employment as missiles aimed at the breaking up of the nuclei of atoms. Perhaps " aimed " is the wrong word, for what is done is to bombard a substance by the particles and hope for lucky hits on the nuclei, which sooner or later are virtually bound to come. Consider a stream of electrons passing through a gas. If their energy is very small, they will have no effect at all. If it is increased somewhat, they disturb the outer electrons of the atoms of the gas and cause emission of light, as in the cathode ray tube. If the energy is still greater they disturb the inner electrons and give rise to X-rays. If it is greater still, they may change the nucleus if they hit it, and then remarkable things happen. The chance that any particular particle will do this is, of course, extremely small, since the nuclei occupy so small a proportion of the space filled by the gas, but out of several million attempts some may be depended on to succeed.

The results may be of several kinds. The bombarding particle may be a positive proton, a doubly positive α-particle, a negative electron, or a neutron. Neutrons are particularly effective, for, being uncharged, they are not deflected by the outer atomic electrons. Even slow neutrons may therefore have great penetrating power and be able to achieve direct

hits on the nuclei. When a hit occurs the bombarding particle may remain in the nucleus, or knock a part of the nucleus away, or disturb the equilibrium of the nucleus so that it changes its configuration and emits a γ-ray. It is impossible yet to predict with confidence what will happen in particular cases, and we have to rely on experiment to tell us. The immediate possible alternatives, however, are clear enough. If the nucleus absorbs a negative charge, its total positive charge is decreased, and it becomes the nucleus of an element of lower atomic number. If, on the other hand, it receives a positive charge, it becomes the nucleus of an element of higher atomic number. We have thus a transmutation of elements—a phenomenon dreamed of by the ancient alchemists, but achieved by methods beyond their imagination.

The Wilson Chamber : The results of the bombardment of atoms are made visible by a most ingenious piece of apparatus devised by C. T. R. Wilson. It has been said (I, 174) that condensation from the vapour to the liquid state requires the presence of nuclei, such as particles of dust, on which drops of liquid can form. It is found that ionized atoms will serve as such nuclei, so that if we have a dust-free space, slightly supersaturated with water vapour, and ions are created in it, the places where the water condenses are the places where the ions are. Now a particle passing through a gas ionizes atoms along its track by knocking off some of the outer electrons (this, of course, occurs much more frequently than an impact on the nucleus, because there are so many more outer electrons). Consequently, a trail of water-drops appears, indicating the path taken by the particle. The energies with which the particles are projected are so much greater than the energy needed to ionize atoms that the particles are not appreciably deflected from their paths by this process, but matters are different when they strike a nucleus. Here they meet particles whose masses are usually greater than their own, and, as a result of the impact, the bombarding particle

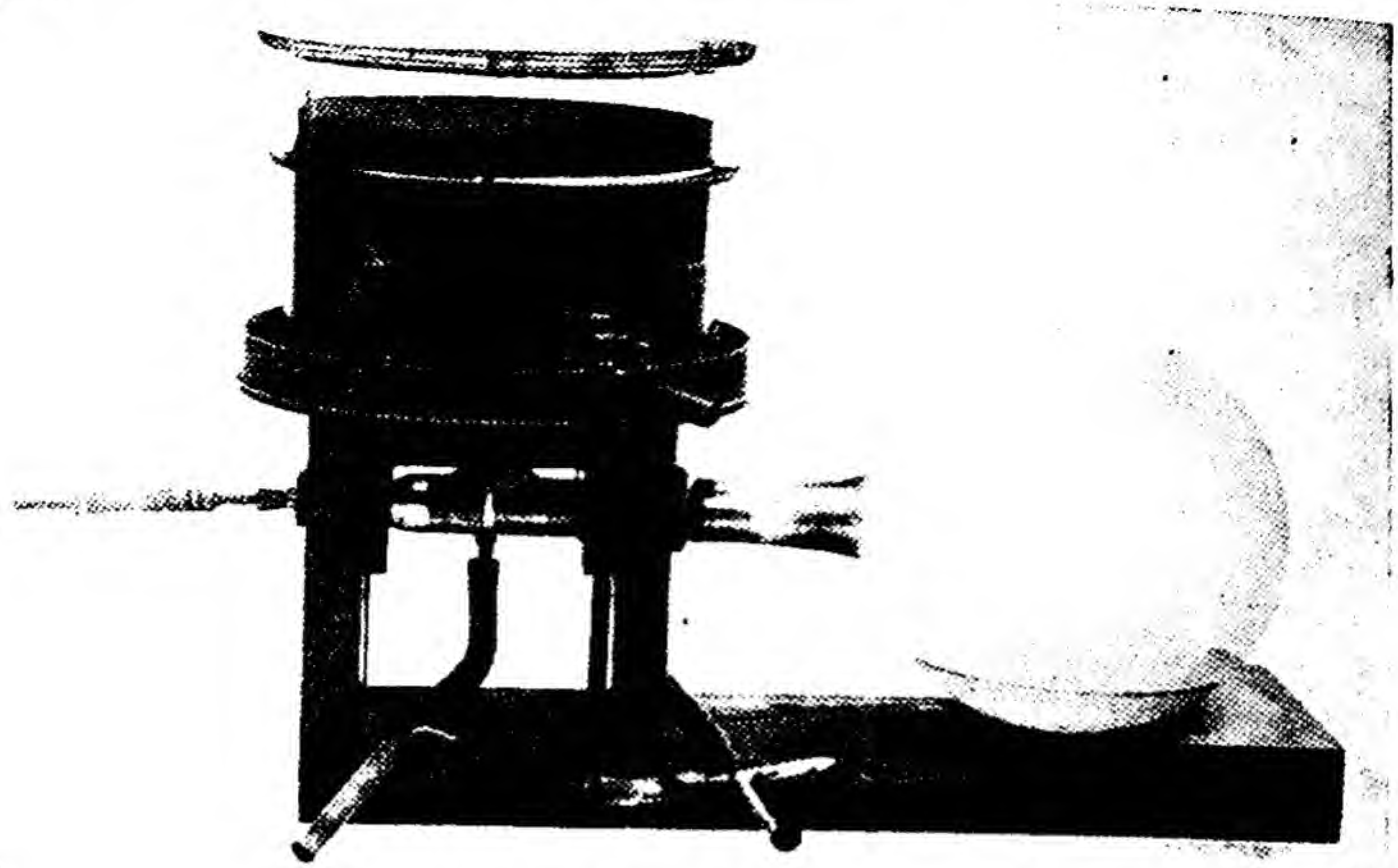


FIG. 143

The Wilson Cloud Chamber

and the nucleus move off at a sharp angle to one another. All this is shown up beautifully by the trail of water drops which they leave behind, and when this is photographed it shows clearly what has happened. The apparatus is very simple: Fig. 143 shows an example constructed in 1911, which embodies all the essential features of instruments now in use. The track of an α-particle is shown in Fig. 144. The particle is moving from left to right, and the track shows first a slight deviation caused by a fairly close approach to an atomic nucleus in the chamber, and shortly afterwards a more direct hit in which the particle is sharply deflected. The beginning of the path taken by the nucleus hit in this second encounter is also shown as a very short line. Many details can be learnt from a careful study of the tracks of particles in the *Wilson chamber*, as the containing vessel is called.

FIG. 144
Track of an α-particle

Artificial Radioactivity and Chain Reactions : Although this work on the bombardment of atomic nuclei is as yet only in its earlier stages, two consequences of the highest importance have already emerged. The first is the discovery of continuous processes of atomic disintegration (or "fission," as it is sometimes called, after the method of reproduction by division characteristic of certain lower forms of life) ; the second is the release of nuclear energy.

Continuous disintegration is of two types. The first is exemplified by an isotope of uranium with an atomic weight of 235 (its symbol is written ^{235}U) when it is bombarded by slow neutrons. This process will be mentioned again shortly in connection with the release of energy which accompanies it, but the point to note here is that when a disintegration occurs, fresh neutrons are among its products, and these act as missiles which can break up other nuclei. A self-perpetuating process, or "chain reaction," is thus started which in favourable circumstances may perhaps continue indefinitely.

The second kind of continuous disintegration follows the production of new radioactive atoms. In the naturally occurring radioactive elements, nuclei break up spontaneously, without bombardment, but these elements are almost entirely confined to the heaviest in the chemist's table. The result of

bombarding lighter elements, however, has been in some instances to form nuclei which are themselves unstable and become radioactive. This phenomenon is known as *artificial radioactivity*. It is already clear that it will play an important part in the future application of physical knowledge to medicine, among other things. Artificially produced radioactive elements can be used, for example, instead of radium in the treatment of cancer and other diseases, and a quite new field of investigation has been opened up by their employment as indicators of the course of elements in their passage through organic bodies. Chemically, radioactive phosphorus, for instance, behaves exactly like ordinary phosphorus, and if medicine containing a very small amount of the radioactive element is injected into the bloodstream, the rate at which it travels through the system can be determined by the great ease with which extremely small quantities of a radioactive substance can be detected in blood extracted from various parts of the body. The rays emitted readily discharge a gold-leaf electroscope, and may be detected by other sensitive instruments also. This is probably the most delicate test known for the presence of minute quantities of an element.

Nuclear Energy: It has for some time been evident on theoretical grounds that a large amount of energy is stored up in the nuclei of atoms, which is potentially available for our use if a suitable means of releasing it can be found. Such a means might be provided by the bombardment experiments we are considering. The assemblage of particles forming an atomic nucleus must possess some potential energy in consequence of its stable configuration, just as the system of outer atomic electrons possesses potential energy (see p. 22). If, then, as the result of a bombardment, this assemblage is broken up or rearranged, some of this energy may be superfluous in the new configuration formed, and may be released in some form such as γ-rays or kinetic energy of the products of disintegration or some other kind of energy

which may be made to do work. It is believed, in fact, that this is the source of the vast amounts of energy continuously radiated into space by the Sun and stars. In the interiors of these bodies, where the temperatures and pressures are extremely high, nuclear bombardment must be naturally and constantly occurring, and a highly probable scheme has been worked out by which, through a series of changes, the hydrogen in a star (the stars are believed on independent grounds to possess a very high proportion of this element) is gradually transformed into helium, with the release of energy which leaves the star as ethereal radiation, including light.

We have not yet copied this particular process on the Earth, but other nuclear changes have been brought about which release energy vastly greater, in proportion to the mass of substance involved, than any terrestrial physical process previously known. The " atomic bomb " is the first large-scale application of this method of producing energy, but it may reasonably be hoped that peaceful applications will be the predominant influence, on the material side, in the future development of our civilization.

To understand the principles involved in this matter the first thing to notice is that Coulomb's law—that the forces between electrified particles vary inversely as the square of the distances between the particles—must break down when those distances are very small. By " very small " we must understand something considerably less than the diameter of an atom (*i.e.* the diameter of an electronic orbit—about 10^{-8} cm.) because we are able to give a satisfactory and detailed account of the process of emission of light by the movement of electrons from one orbit to another (see p. 22) on the assumption that the nucleus attracts the electrons according to Coulomb's law. There is strong reason to believe, however, that if an electron could get inside its smallest orbit and approach the nucleus, the attraction would diminish and a point would be reached where it would disappear altogether, and begin to change to a repulsion.

That some such change of law must occur on very close approach is suggested by the most ordinary and commonplace experience. Consider a rod of iron or any other solid substance. It consists of atoms which are practically "in contact," *i.e.* the orbits of their electrons as nearly as possible touch one another, so that the electrons themselves must frequently pass at extremely close quarters. Now we know that we must exert a very great force to compress the rod, so we must conclude that a closer approach of the atoms is resisted by a strong repulsive force. But equally we must exert a very great force to extend the rod, so we must conclude also that a separation of the atoms is resisted by a strong attractive force. The atoms are therefore in a configuration of great stability, from which it is extremely difficult to displace them in any direction. Such a state would be impossible if Coulomb's law held at all distances between the particles.

The same conclusion follows from a consideration of the equilibrium of the nucleus. Here we have a collection of protons and neutrons separated by distances estimated to be of the order of 10^{-13} cm. or less. The protons are all positive charges and the neutrons have no charge at all, so that if the forces between the particles obeyed Coulomb's law, an atomic nucleus, if it could be once formed, would immediately fly apart with such energy that everything in the neighbourhood would be destroyed. Let us calculate, for example, the energy required to bring a proton within 10^{-13} cm. of another proton ; this energy, of course, would be released again if the particles separated. The nucleus of a helium atom contains two protons, and what we are about to evaluate may be thought of as the energy of formation of a nucleus of this element. The charge of a proton is 5×10^{-10} E.S.U. (p. 135), and the potential at a distance of 10^{-13} cm. from it—*i.e.* the work done in bringing unit charge to this point from infinity—is therefore $\frac{5 \times 10^{-10}}{10^{-13}} = 5000$ ergs per unit charge. The

work done in bringing up another proton is therefore 5000 $\times 5 \times 10^{-10} = 2{\cdot}5 \times 10^{-6}$ ergs. Now in 1 cc. of helium at normal temperature and pressure there are about $2{\cdot}7 \times 10^{19}$ atoms. Hence the energy which would be released from the nuclei in 1 cc. of helium if Coulomb's law suddenly began to operate would be of the order of 7×10^{13} ergs, or one and a half million calories. This energy, of course, is not released, so the forces between protons in the nucleus must be very different from those between large-scale electrified bodies. We do not know exactly what they are, but it is clear that if the nucleus were disturbed so that the particles were separated to distances at which the ordinary repulsion could operate we should have an enormous release of energy. The nuclei of most atoms contain many more than two protons, so the release of energy on this account would be greater, but we must not attach too precise a significance to such calculations as that just given. The problem has been simplified by the neglect of certain factors—*e.g.* the influence of the atomic electrons—and only the general order of magnitude of the energy to be released can be arrived at in this way.

The disintegration of ^{235}U, already mentioned as an example of a chain reaction, is remarkable also because of the great amount of energy released per atom. In most of the disintegrations so far achieved the change is slight ; the bombarding particle adheres to the nucleus or else knocks off a small part such as a proton or, at most, an α-particle. But shortly before the latest war began it was discovered by Hahn that an isotope of uranium, later found to be ^{235}U, broke into particles of roughly equal mass, and the energy released per atom was much greater than that accompanying the ordinary smaller changes. The precise course of the changes is perhaps not yet completely known. Barium and lanthanum, with atomic numbers 56 and 57 respectively, were first observed, but it later became probable that these arose as secondary products of artificial radioactive isotopes of xenon and caesium (atomic numbers, 54 and 55), and since

uranium has an atomic number 92, something or some things having a total atomic number of the order of 37 or 38 must have appeared also ; strontium (atomic number 38) was, in fact, soon detected. Whatever the sequence of changes, however, it was certain that an abnormally large amount of available energy was released in the process, and since, as soon appeared probable, it was accompanied by the emission of neutrons which could produce further such changes and so initiate a chain reaction of the kind already described, the prospect of making practical use of nuclear energy first took definite shape almost at the moment when Hitler invaded Poland.

It was inevitable in these circumstances that the idea of using such energy for destructive purposes should claim the attention of those responsible for the scientific side of the war effort, but it cannot be emphasized too strongly that the original impetus which led both to the discovery and to its further development was simply the desire for knowledge. Hahn made and announced his discovery in Germany, and so late as July 1940 a letter dated 3rd May was published in the English journal, *Nature*, from five Japanese scientists, reporting the results of their work on the fission of uranium by neutron bombardment, which was carried out as a part of the programme of the Atomic Nucleus Sub-Committee of the Japan Society for the Promotion of Scientific Research. Like all fundamental scientific work the investigation was international, and its results were published for the information of all and sundry, and we may be sure that when war dominated the thoughts of nations, every country concerned intensified its efforts to make an " atomic bomb " a practical possibility, and began to impose the strictest secrecy on its scientific workers.

The problems they had to solve are simple enough to state. A sufficient supply of uranium had first to be obtained, a practicable method had to be found for separating the effective isotope (only a very small fraction of the whole)

from the other isotopes of uranium, and a mechanism devised for initiating the reaction at the time and place required. Granted a sufficient concentration of ^{235}U, the energy with which the products of each disintegrated atom would fly apart would amount, even with a bomb small by existing standards, to a destructive agency enormously more powerful than anything previously used. Fortunately the United Nations had access to the lion's share of the raw material, and were able to spend sufficient money on the task of separation. Perhaps it is not inaccurate to say that they had at their service the best scientific minds also, though with the knowledge available in 1939 any of the enemy countries could easily have commanded the necessary intelligence if the other requisites had been forthcoming. It was a race, and we cannot be too thankful that it was we and our allies who won.

EXERCISES

1. How are electrical oscillations in space produced? State how you would arrange a circuit to give oscillations of a prescribed frequency.

2. Describe and explain the phenomena occurring when a discharge passes through a gas whose pressure is gradually reduced. What differences would you expect to observe between the discharges through different gases?

3. Give an account of the method of production of X-rays. What atomic processes are believed to be responsible for the emission of these rays, and to what uses may the rays be put?

4. What is photo-electricity? How is it related to the absorption of light and to the atomic structure of the elements?

5. Describe the thermionic valve, and state some of its applications.

6. Write a short essay on the transmutation of the elements.

CHAPTER XV

TERRESTRIAL MAGNETISM AND ELECTRICITY

Terrestrial Magnetism

Uniformly Magnetized Sphere

It has already been mentioned that the Earth acts as though it were a magnet with poles in the neighbourhood of the geographical poles. We have not specifically considered

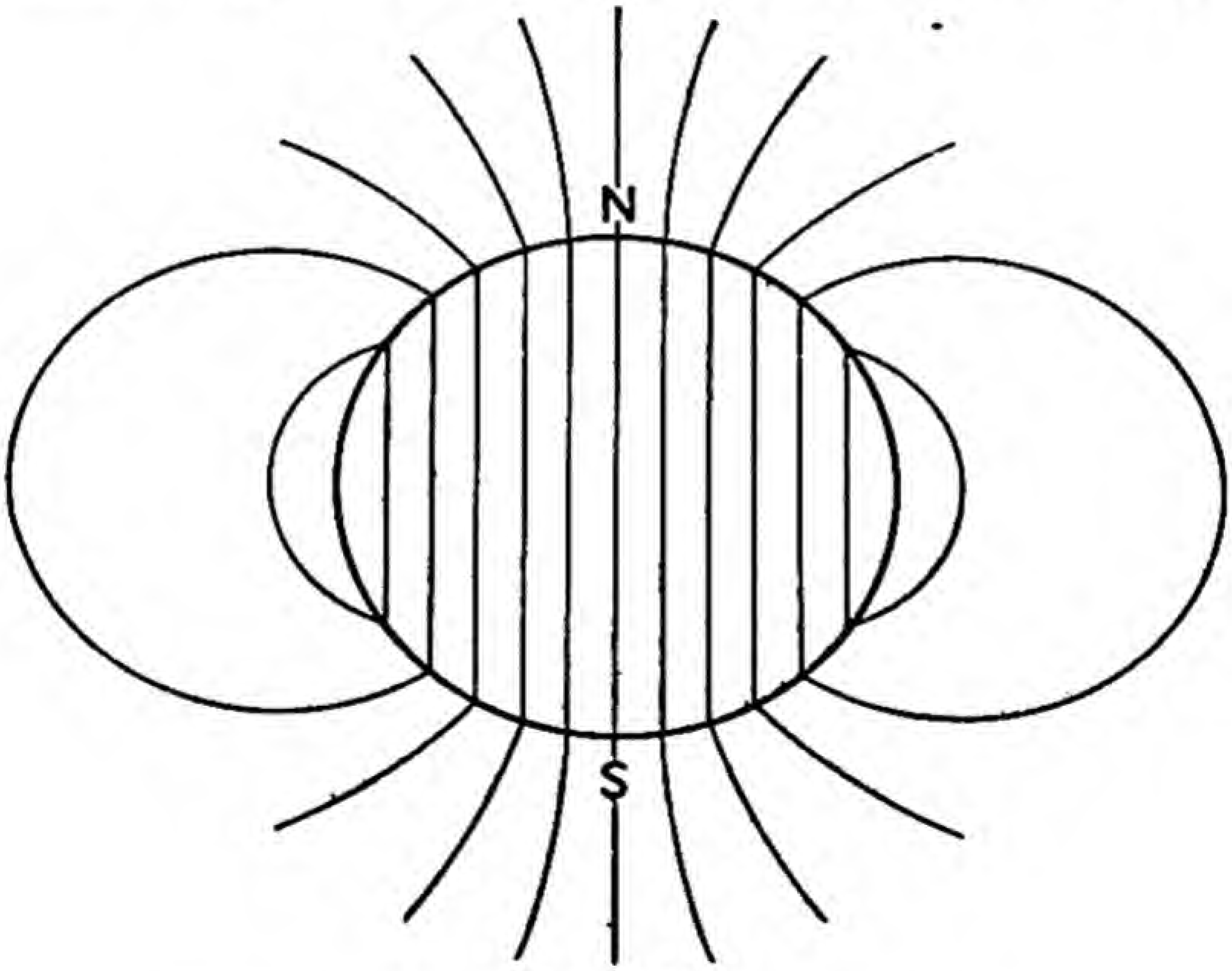


Fig. 145
Field of Uniformly Magnetized Sphere

spherical magnets in this book, since the mathematics involved is rather too advanced, but the student will have little difficulty in understanding in general terms what is meant by a

uniformly magnetized sphere. One hemisphere acts as a N. pole and the other as a S. pole, and the surrounding field has its regions of greatest intensity near the central points of the two hemispherical surfaces, which are therefore often referred to as the poles. In terms of lines of force the field is as illustrated in Fig. 145. The lines inside the sphere are crowded together, and are straight and parallel, while those outside are in general form similar to those around a bar magnet.

The Regular and Irregular Terrestrial Field

When the Earth's external field is mapped, it is found to approximate closely to this, but not to be quite identical with it. We can, therefore, regard it as being made up of two parts—a *regular* field, which is that of a uniformly magnetized sphere, and forms at least 94 per cent. of the total field, and an *irregular* field, which is a relatively very weak field superposed on the former.

The Magnetic Elements

The mapping of the Earth's field is a very laborious task, involving observations at widely scattered places, including many at sea by a specially constructed non-magnetic ship. The field at a given place is generally inclined to the Earth's surface, and the two components (namely, those parallel and perpendicular to the surface, known as H and V, respectively) are determined separately. The inclination θ of the resultant field to the ground is known as the *angle of dip*. Clearly

$$\frac{V}{H} = \tan\theta \qquad . \quad . \quad . \quad . \quad . \quad (15{\cdot}1)$$

If the Earth's magnetic poles were coincident with its geographical poles, the direction of H would always be due north and south. Actually it is somewhat different from this. The angle between the two directions—between geographical north and magnetic north, for instance—is called

the *declination*, a term which must not be confused with the declination of a point on the celestial sphere (I, 28).

The distribution of magnetic force over the Earth is represented in maps by lines joining places of equal magnetic

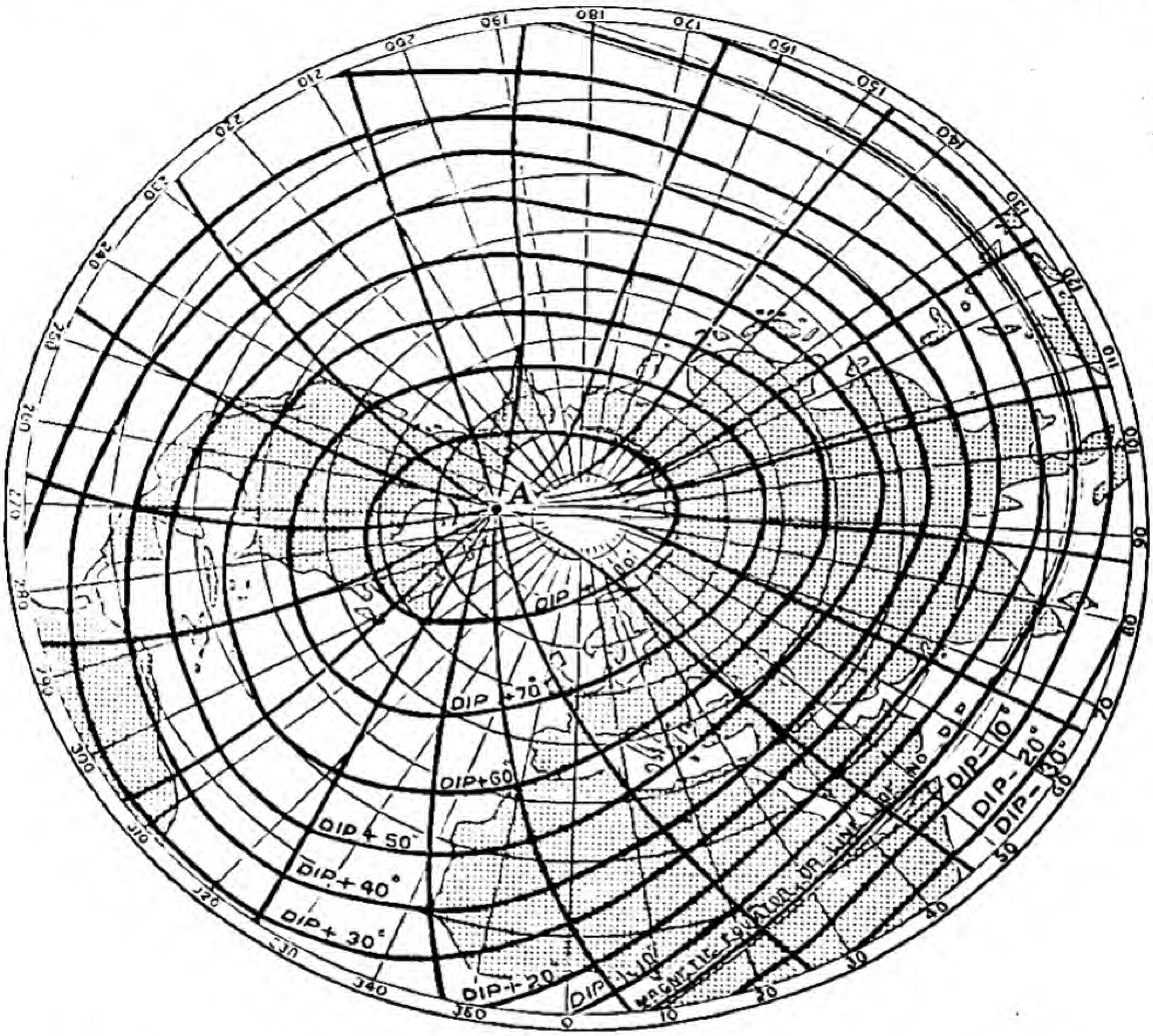


FIG. 146

Map of Northern Hemisphere showing Horizontal Magnetic Component (H) and Lines of Equal Dip (about 1830)

force (*isodynamic* lines), equal dip (*isoclinic* lines), and equal declination (*isogonic* lines). The horizontal or vertical component, H or V, is often plotted instead of the total force, giving lines of equal horizontal or vertical force which have been given no special name. Figs. 146, 147, and 148 show the

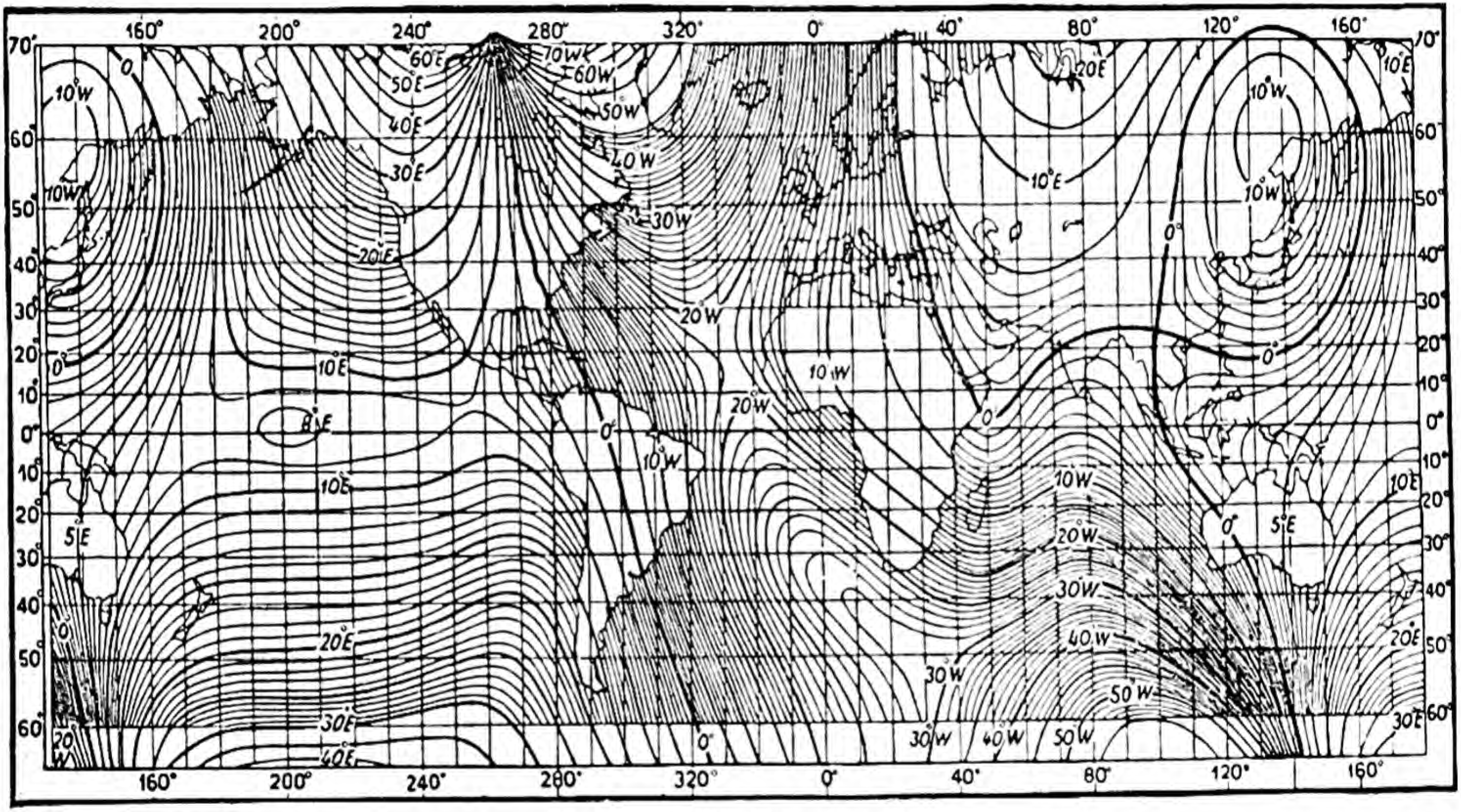


Fig. 147

Admiralty Chart of Isogonic Lines for 1922

courses taken by some of these lines for certain epochs. Since they change continuously, periodic re-determinations of the *magnetic elements*, as these variables are called, have to be made.

Variations in the Magnetic Elements: The magnetic elements would, of course, vary from place to place even if the Earth's field were perfectly regular, for, as Fig. 145 shows, the lines of force meeting the surface are not equally densely crowded, or equally inclined, at all places. In addition to these variations, however, there are disturbances due to local conditions. A deposit of magnetic material, for instance, or

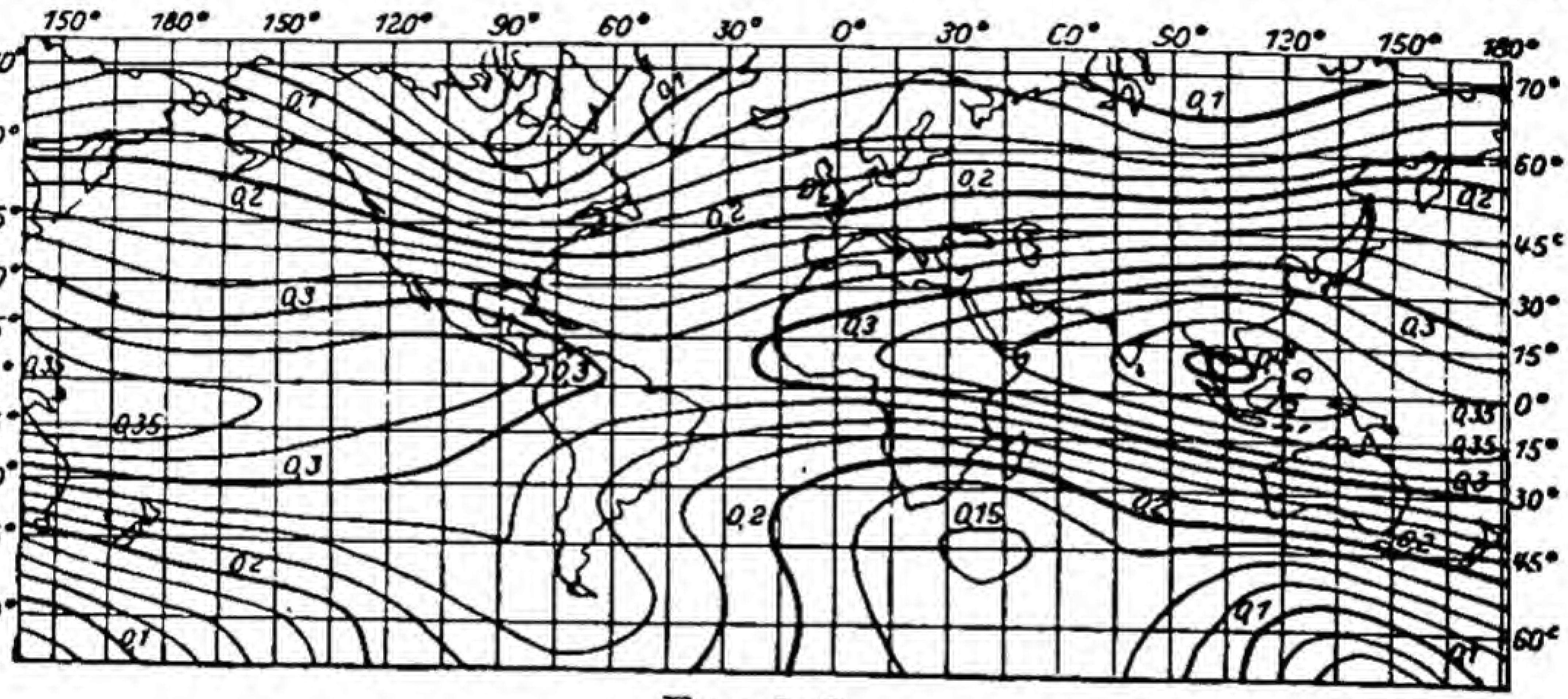


FIG. 148

Lines of equal Horizontal Intensity (H), 1922

the neighbourhood of volcanoes, causes considerable local disturbance of the field. It is believed that the lava from volcanoes becomes magnetized on solidifying, and remains as permanent magnetic material.

The elements vary also with time. These changes are rather complex. In addition to small diurnal and annual changes, there appear to be certain secular (*i.e.* slow continuous) changes, though there is not enough evidence of their trend to enable us to predict future values with much certainty. In Fig. 149 are shown the changes in both declination and dip (plotted as " inclination," a name by which dip

is sometimes known) at London between 1540 and 1920. The part of the course before 1580 is speculative, since the existence of dip was not discovered until 1576. The course suggests a periodic change, the period being about 480 years. Observations at other places, however, give different periods, and, in fact, the existence of any regular period is very doubtful.

The intensity of the Earth's field, besides suffering small casual changes, appears to be steadily decreasing with time. The intensity is generally measured in terms of the electro-

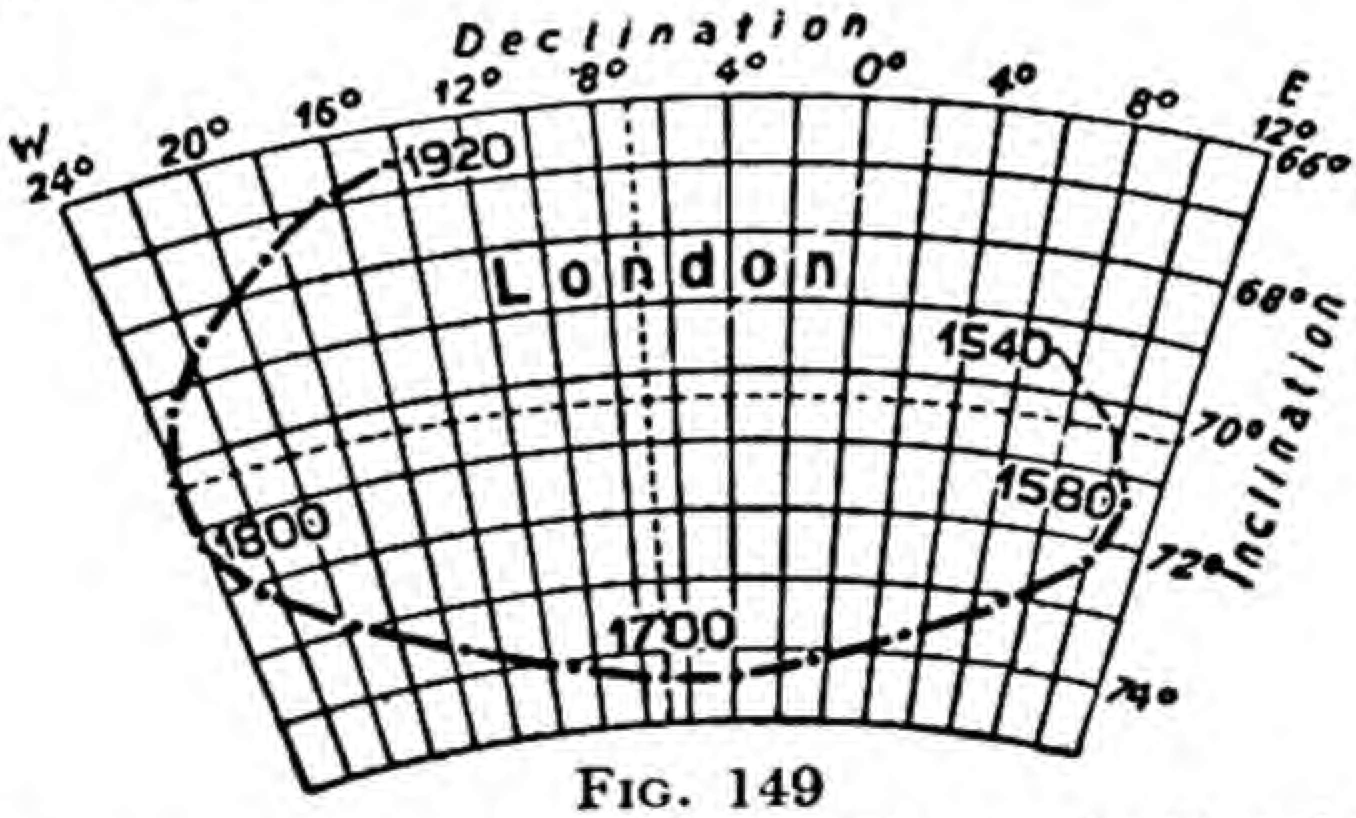


FIG. 149

Changes in Declination and Dip (or Inclination) at London between 1540 and 1920

magnetic unit known as the *gauss* ; it is the intensity of a field in which unit magnetic pole is acted on by a force of one dyne. The intensity of H_0, the horizontal component of the *regular* field, appears to have decreased from about 0·327 gauss in 1829 to 0·316 gauss in 1922. Though this change may appear small, yet compared with the usual scale of terrestrial changes it is a very rapid process. We have not yet sufficient data to determine whether it is continuously in the same direction or is part of a cyclic change.

The Origin of the Earth's Magnetic Field

The origin of the Earth's magnetic field, as well as that of its regular variations, is almost entirely unknown. The close

approximation of the field to that of a uniformly magnetized sphere suggests that the regular part is of internal origin. The irregular remainder may be due in part to electric currents passing from the ground to the air or vice versa ; but though we know that such currents exist, they appear to be too weak to account for more than a fraction of the irregular field.

The curve of fluctuation of the field follows closely the curve showing the variation with time of the number of spots on the Sun, which has a maximum every $11\frac{1}{3}$ years, with, of course, intervening minima. It is believed that a sunspot is a source of emission of electrons and ionized atoms which travel outwards and of which some enter the Earth's atmosphere. Being in effect electric currents, streams of such particles have magnetic fields associated with them, which cause changes in the Earth's field. The incidence of these particles on the gases of the upper atmosphere displaces electrons from the atoms of those gases, with a resulting emission of light in the manner described earlier. This light constitutes the *aurora*—a phenomenon seen best near the Earth's magnetic poles, since the particles from the Sun are deflected towards those regions by the Earth's field.

Applications of Terrestrial Magnetism

The earliest, and still the chief, application of terrestrial magnetism to practical uses is in navigation. The *mariner's compass* consists of a magnetic needle, usually of a special form, freely pivoted over a graduated card. Provided that the needle is protected from all magnetic influences other than that of the Earth, the needle will always point towards one of the Earth's magnetic poles. The bearing of one point on the horizon is thus indicated, and so the azimuth of any other point is known. The compass is now used in aerial as well as in marine navigation.

Another application—concerned with local variations rather than with the main field—is to prospecting. Irregularities in the magnetic properties of materials under the

Earth's surface show themselves in irregularities of the field near the surface, and experience enables us to deduce therefrom a good deal about the structure and composition of the subterranean regions. The method is much used in the search for minerals or oilfields.

The variations of the Earth's field have important influences on telegraphy and radio. When violent changes take place—generally as a result of particularly active sunspots—*magnetic storms* are said to occur, and both wireless and cable telegraphy may be seriously affected.

Atmospheric Electricity

The Atmospheric Electric Field

It has been found that there is a permanent electric field in the Earth's atmosphere ; the surface of the Earth is negatively charged, and the upper air positively. This may be shown by connecting two insulated conductors, placed at different heights, to opposite quadrants of a quadrant electrometer, when the difference of potential can be observed and measured.

The rate of change of potential as we go upwards—represented by $\frac{dV}{dh}$, where V is the potential and h the height above the ground—is numerically equal to the strength of the field (see p. 146). Its value varies considerably with weather conditions ; it is greatest in fine weather, but at all times there are wide variations from place to place on the Earth's surface, and from height to height above it. The average value near the ground is about 100 — 120 volts per metre, which seems extremely great. The Earth is so large, however, that the negative charge necessary to produce this field is only $2{\cdot}7 \times 10^{-4}$ e.s. units (9×10^{-14} coulombs) per sq. cm., so that we are quite justified in ordinary considerations in regarding the Earth's surface as uncharged. The variations of the field near the ground are exemplified by the values at Kew and at Davos (in Switzerland), which are 317 and

64 volts per metre, respectively. Over the sea the field is approximately independent of geographical position, and has a value of 126 volts per metre. Furthermore, on land, fluctuations with time, even at the same place, are very large, amounting perhaps to 50 per cent. of the average value. These are caused largely by local pollution of the atmosphere, and are partly periodic. Thus there is a diurnal variation, with maxima and minima occurring at the same times (19h. and 5h. G.M.T. respectively) all over the Earth, and on this are superposed local periodic variations. There is also an annual variation, with a maximum in the local winter and a minimum in the local summer, except in the Antarctic, where the maximum is in the summer and the minimum in the winter.

These data are for conditions near the ground. As we ascend in the atmosphere the field decreases very rapidly. At a height of 10 km., for instance, the field is less than $\frac{1}{50}$ of the ground value. At 15 km. the P.D. between the air and the ground is about 300,000 volts, and above this height the conductivity of the air is so great (owing, no doubt, to electrons liberated by photo-ionization of the atmospheric gases by sunlight) that the potential may be regarded as constant.

The Atmosphere as a Condenser: We may thus regard the Earth's surface and the upper atmosphere as two plates of a condenser, with the intervening air as the dielectric. If this dielectric were a perfect insulator there would be no difficulty in accounting for the maintenance of the atmospheric field, for there would be no means for the condenser to discharge itself. It is, however, anything but a perfect insulator; as we shall see in a moment, it contains many charged ions, the motion of which under the influence of the field, assisted by convection currents and falling rain and snow, would tend to discharge the condenser in a time estimated to be about ten minutes. The problem therefore arises: how is the atmospheric field maintained? No final answer

has yet been given to this question. It appears, however, that contributory factors are the conveyance of electric charge to the ground by lightning discharges (of which, taking the Earth as a whole, about twenty are estimated to take place between thunderclouds and the Earth each second), and the discharge of electricity from points or relatively sharp places on the Earth, such as trees. We have already seen (p. 134) that the density of charge on a surface is greatest where the curvature is greatest, and when it reaches a certain value, ions or electrons spontaneously leave the surface. This effect seems to have important consequences in nature.

Ions in the Earth's Atmosphere

The gases in the Earth's atmosphere are continually being ionized through several causes. Of these the chief are the particles discharged from radioactive matter in the Earth's crust and in the air itself, and cosmic rays. The energy of the particles from these sources is able to remove electrons from the atmospheric atoms or molecules, thus creating free positively and negatively charged particles. The direct influence of sunlight is mainly confined to the upper layers of the atmosphere, the short ultra-violet waves necessary for the photo-ionization of the gases having used up their energy before they can penetrate very far towards the lower layers.

The charges formed by the ionization of the lower atmosphere are known as *small* ions. Over land most of these attach themselves to nuclei such as bits of dust, drops of water vapour, etc., forming *large* ions which outnumber the small ions by about ten to one. Over the sea, however, where the air is purer and more constant, the small ions predominate. In these regions cosmic rays are practically the only effective ionizing agency.

Atmospheric Currents

The existence of these ions in the atmospheric electric field results in an electric current passing between the ground and

the upper air. The ions move in direct obedience to the field, and are carried also by convection currents. It is found that the mean value of the fine weather current is about 2×10^{-16} amperes per sq. cm. The part of this current arising from direct conduction (apart from convection) is much more constant than the electric field—the resistance of the air seems to vary with the field so as to maintain a fairly constant current according to Ohm's law.

The resistance of the air may be considered as made up of two parts—that of the air near the ground and that of the upper air (below the level of 15 km., which we have regarded as the upper plate of the atmospheric " condenser "). Most of the variations are in the former, owing to local fluctuations, so that the air is, in effect, a constant upper resistance and a smaller variable lower one connected in series. The local, transient effects of rain and snowstorms are very great, and may quite swamp the permanent fair weather conditions, even the direction of the field being sometimes reversed. Single drops of rain or snow, for example, may carry charges of 100 or 200 e.s. units, and their potentials may reach 300 volts.

Lightning

Lightning is due to the discharge of electricity between two oppositely charged clouds, or between a charged cloud and the ground, when the P.D. is so great as to break down the resistance of the intervening air. This occurs when the velocity of the comparatively few ions in the air, under the influence of a large field, becomes so great as to ionize large numbers of atoms with which the ions collide. The many ions thus produced make the air a conductor, and a discharge passes with great violence. The potential gradient necessary for breakdown of the resistance of the air is about 10,000 volts per cm., and the P.D. between two thunderclouds when a discharge passes is of the order of 10^9 volts.

The origin of the electrification of clouds is internal, but

very little is yet known with certainty about its mechanism. The effect, however, is that positive charges go to one side and negative charges to the other, so that the cloud acquires a *polarity*. When it discharges, the same causes soon restore this polarity, so that a series of discharges occurs, an active cloud taking part in a lightning flash about every 20 seconds.

Lightning Conductors : When a discharge takes place between a cloud and the Earth, its path has the well-known zigzag form, which is that of least electrical resistance. Where it meets the Earth the discharge may cause fire or other damage, and to prevent this occurring to large buildings lightning conductors are used. A lightning conductor consists of a straight conducting rod, whose lower end makes good electrical contact with the Earth, and whose upper end is pointed and reaches above the highest point of the building. Its action is twofold. In the first place, by the discharge of electricity from its pointed end it lowers the potential gradient (the electric field) in the neighbourhood, and so makes the discharge tend to occur elsewhere where the field is stronger. Secondly, if the discharge should nevertheless reach the building, it chooses the conductor as the easiest path, and so passes harmlessly down into the Earth without damaging the building.

EXERCISES

1. Describe the chief phenomena of terrestrial magnetism. What experiments would you make to measure the horizontal component of the Earth's field ?

2. Give an account of the electric field existing in the Earth's atmosphere.

3. How are ions produced in the Earth's atmosphere ? Describe the effects of the existence of these ions in the atmospheric electric field.

INDEX

NASEER ZARGAR
M.Sc. (Final)

PRINTED IN GREAT BRITAIN AT
THE PRESS OF THE PUBLISHERS

Borrower's No.	Issue Date	Borrower's No.	Issue Date
